Sublexical representations in auditory word recognition: evidence from lexical learning

Lukas Wiget

PhD Thesis
The University of Edinburgh
2007

Abstract

The main question this thesis addresses is whether auditory word recognition proceeds directly from the input to the lexicon or whether there is a prelexical level of processing where segmental units are recognised.

In the first part, I situate the question in a wider context of representational issues, and show that it is a crucial question because it allows us to distinguish two broad types of word recognition models: what may be called *direct-* and *mediated-access* models. A review of the research literature addressing this question shows that existing experimental results are inconclusive.

The second, experimental, part of the thesis addresses the research question with a lexical learning paradigm. English-speaking subjects are first trained to recognise novel words that contain a non-native speech sound (a voiceless bilabial fricative); they then perform two tasks designed to determine whether they have acquired prelexical representations for the non-native segment. The tests used are repetition priming and phonetic categorisation.

The results of the repetition priming task are consistent with direct-access models; but for methodological reasons they have to be regarded as inconclusive. The results of the phonetic categorisation task favour mediated-access models. They also suggest that the representations used at the prelexical level of processing are more likely to be position-specific segmental representations rather than syllable rhymes.

These results are compared with those of other studies. They are consistent with a growing body of evidence that auditory word recognition involves a prelexical level of processing where segmental representations are recognised.

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Lukas Wiget

Acknowledgements

In carrying out the research that lead to this thesis, I have profited greatly from the help, advice and encouragement of my two supervisors Alice Turk and Mits Ota. My greatest thanks must go to my first supervisor, Alice, who always gave me sound practical advice and the benefit or her critical eye (and ear), while leaving me enough space to develop my own ideas.

I also wish to thank everyone in the Department of Linguistics and English Language (and the former Department of Theoretical and Applied Linguistics), particularly its friendly and helpful computing staff, the members of the Phonetics and Phonology research group, all fellow common room regulars and my office mates. My thanks also go to all of you who have made my time in Edinburgh such a rich and pleasant experience, and to my friends back home who I, unfortunately, neglected somewhat in the past few years.

I should also thank the Janggen-Pöhn-Stiftung, the Swiss National Science Foundation and the University of Edinburgh for their sponsorship of my PhD. Finally, I wish to thank my parents for their unfailing love and support.

Contents

Introduction

Auditory word recognition is the study of how words are recognised from the acoustic speech signal. The recognition of words is an important and probably necessary step in the comprehension of spoken language. It is a common assumption that human listeners possess a *mental lexicon* were words are stored in an accessible form, together with the syntactic, semantic and pragmatic information necessary to comprehend what is being said by the interlocutor. Questions arise about the form of these lexical representations, and about whether words are recognised directly from the acoustic signal or whether smaller, sublexical units are required for word recognition to be possible.

The present thesis asks the following four questions about the lexicon and the process of word recognition. The second question is the one I will address experimentally.

1) Are lexical representations (i.e. the words in the lexicon) unstructured wholes, or are they composed of smaller, sublexical representations (such as syllables, phonemes or distinctive features)?

2) Does the speech signal map directly onto these lexical representations, or are there prelexical levels of processing?

3) If a prelexical level of processing exists, do the representations used at the prelexical level correspond to the sublexical representations used in the lexicon?

4) How abstract are the representations used at any given level of processing?

In Part I, I will further elaborate and explain these questions (Chapter 1 and Chapter 2), show why I think question 2 regarding the existence of a prelexical level of processing deserves our greatest attention (Chapter 3), and consider what the research literature has to say about it (Chapter 4). I will conclude that the question whether word recognition is direct or indirect has not yet been answered sufficiently, and that additional research is therefore needed.

Part II reports my attempt to undertake this additional research. The method used was a word learning paradigm, where English-speaking subjects were made to acquire novel words that contained the non-English voiceless bilabial fricative /ɸ/, and were then tested on whether

they had formed prelexical representations for the non-native sound. Chapter 5 describes the general design of the experiment. Chapter 6 reviews some previous studies that have used the tasks chosen as my test tasks (repetition priming and phonetic categorisation); this is meant to justify and elucidate my use of these tasks. Chapter 7 then describes how the design was implemented, and Chapter 8 states the predictions about the test tasks. Chapters 9 to 11 finally present the outcome of the experiment, separately for the training task and the two test tasks.

Part III is a discussion of the outcomes of the experiment. I will first discuss the results on their own terms, and then relate them to the findings of other studies. I conclude that auditory word recognition is more likely to be indirect than direct: the process of word recognition seems to require a prelexical level of processing where units smaller than words are recognised. I will further conclude that these pre- and sublexical units appear to be segments; and they are unlikely to be phonemes but rather position-specific allophones.

Part I

The research question

In the four chapters of Part I, I will develop the research question sketched in the introduction, and then discuss the relevant research literature.

Chapter 1 reviews the major models of word recognition in order situate the present study in a wider context, and to provide a basis for the theoretical discussion that follows in the next two chapters. Chapter 2 describes the main representational dimensions along which models of auditory word recognition can be classified. Chapter 3 presents typology of possible word recognition models, and then narrows it down to the ones that are theoretically well-founded. We will see that the research question of this thesis, namely whether auditory word recognition is direct or indirect, has the potential to distinguish between the main model types. Chapter 4 reviews relevant studies that have addressed the question of the direct or indirect nature of auditory word recognition. The review suggests that additional research is justified – particularly if it employs a different experimental paradigm.

1 / Models of auditory word recognition

The purpose of this chapter is twofold. My first aim is to give a brief overview of common models of word recognition and speech perception. This overview is neither comprehensive nor very detailed; it focuses on the representations that are used in the models, and its purpose is to give a flavour of the issues involved, so that the next two chapters will be easier to follow. The second aim is to motivate the research presented in this thesis – i.e. determining whether auditory word recognition is direct or indirect – in the context of models of word recognition.

Readers who are familiar with the models introduced, or who are convinced of the importance of the question may wish to proceed directly to Chapter 2. The models considered are the word recognition models Cohort, Trace, and briefly Shortlist and PARSYN (§1.2.1); the Motor Theory, Direct Realism, and auditory theories of speech perception (§1.2.2); Klatt's LAFS, Hintzman's MINERVA 2, and Kirsner et al.'s (1987) record-based model of word recognition (§1.3); and, to provide a general framework, Pisoni and Luce's (1987) sequential linguistic account of word recognition (§1.1).

1.1 A sequential account of word recognition

Pisoni and Luce (1987) suggest that auditory word recognition can be broken up into four main stages of processing: (1) auditory, (2) phonetic, (3) phonological, and (4) a higher-order stage. Apart from the initial auditory stage, these stages of processing closely correspond to the representations used in the linguistic analysis of language; we can therefore also regard Pisoni and Luce's discussion as providing a *linguistic* account of auditory word recognition.[1]

Auditory processing. Auditory processing is obligatory and not specific to spoken language. It has two components: peripheral auditory analysis and central auditory analysis. Peripheral auditory analysis takes place in the cochlea and the auditory nerves, and it produces either

[1] See Studdert-Kennedy (1974, 1976) for similar suggestions.

a neuro-acoustic or psychoacoustic representation of the incoming speech signal (or some combination of both). Central auditory processing is assumed to extract more specific pieces of information from the peripheral representation – such as its spectral structure, fundamental frequency, intensity, etc. – and to pass them on to short-term sensory memory. Pisoni and Luce call the pieces of information extracted *speech cues*.

Phonetic processing. This is the first specifically *linguistic* stage of processing. Pisoni and Luce suggest that at this stage, speech cues are mapped onto a set of *phonetic features* which are grouped into feature bundles that represent *phonetic segments*. Several speech cues typically map onto one phonetic feature, for example the spectral distribution of the release burst of a stop consonant and the shape of the formant transition of any preceding or following vowel are both cues to the phonetic feature 'place of articulation'. As this example illustrates, the speech cues of one phonetic feature can occur at different time points in the auditory representation of the speech signal. Phonetic features, and by implication phonetic segments, thus need to be abstract. For segmental representations this means that a sequence of phonetic segments does not correspond to linearly ordered and non-overlapping stretches of the speech signal. It is at this stage of processing that the problem of lack of invariance mentioned in the introduction is addressed in Pisoni and Luce's account. More about this later in §1.2.2.

Phonological processing. At this stage, phonetic segments and features are transformed into phonological segments, i.e. phonemes. This entails that all predictable allophonic variation is discarded: phonetic features that can be predicted from the presence of other features are deleted. In the case of English stop consonants, the feature [± aspirated] may be left out at this stage because it can be predicted from the feature [± voice] and the position in the syllable. Pisoni and Luce further suggest that the output of this processing component may not just be a linearly ordered sequence of phonemes, but may be hierarchically ordered to form syllables.

Higher-order processing. In Pisoni and Luce's framework, higher-order processing refers to any processing *at and above* the lexical level. Two things have to be done at this stage: (i) the mapping of phonological representations onto lexical representations (word recognition), and (ii) the retrieval of the information associated with the recognised lexical item (lexical access) for further processing by the higher-level language processing mechanisms.

This account of auditory word recognition by Pisoni and Luce (1987) is unusual in that it tries to be explicit about the whole process of word recognition from the acoustic signal to the lexicon and beyond. Most accounts either concentrate on the problem of lack of invariance (models of

speech perception), or start with a phonological representation and focus on lexical processing (models of word recognition). We will look at these two types first (§1.2), before considering alternatives that, contrary to the sequential account, present a storage solution instead of a processing solution to the problem of lack of invariance (§1.3).

1.2 Mediated lexical access

1.2.1 The top half: word recognition models

We will first look at the Cohort model, because it is the earliest model that was targeted specifically at auditory word recognition; and then at Trace, because it is the first and most influential connectionist model and thus established a new state of the art in word recognition modelling.

Cohort

Marslen-Wilson's Cohort model (Marslen-Wilson and Welsh, 1978, Marslen-Wilson and Tyler, 1980) was the first word recognition model that was proposed specifically with *auditory word recognition* in mind, as opposed to the recognition of written words or word recognition in general.

The Cohort model assumes that access to the lexicon occurs as early as possible. When a word stimulus is being perceived, all lexical entries that match the initial part of the stimulus (the word-initial cohort) are activated in parallel. As more of the stimulus is heard, the word-initial cohort is winnowed accordingly until a single candidate remains; in which case the word has been recognised and lexical access (i.e. the retrieval of information related to the the recognised word) begins.[2]

The winnowing process does not only take account of bottom-up information (from the speech signal) but also makes use of all the available top-down information. So if a word candidate does not fit the syntactic, semantic or pragmatic requirements that its position in the utterance demand – e.g. if the candidate is an noun when a verb is required – it will likewise drop from the cohort. Top-down reduction of the cohort explains why words can be recognised before they become acoustically unique.

With regard to representational questions the original Cohort model was deliberately non-committal. Its main focus was the time course of auditory word recognition. Nevertheless, for practical purposes at least, representational assumptions had to be made, and Marslen-Wilson and Welsh (1978, p. 56) suggest that a cohort is determined by the initial 150–200 ms of the

[2] In more recent incarnations of the Cohort model, the activation level of candidates that stop matching the input gets reduced, and words are recognised by comparing the activation levels of the two most highly activated candidates (Marslen-Wilson, 1987, 1990).

input and later even speak of the 'initial segment of the incoming word' (p. 60). We can therefore agree with Pisoni and Luce (1987, p. 41), who claim that the Cohort model assumes that both input and segmental representations are based on phonemic units.

Trace

The earliest connectionist model of auditory word recognition, Trace (McClelland and Elman 1986; also Elman and McClelland 1984), has three layers of representations corresponding to features, phonemes and words. The units, or nodes, on each layer can be understood to form hypotheses about the input: phoneme units represent hypotheses about the segment currently processed, word units about the word, etc. A unit on each of the layers is linked to all other units on the same layer by mutually inhibitory connections: e.g. the /p/ unit on the phonemic layer will inhibit the /t/ unit on the same layer, and so on. Units on different layers that are consistent with each other are connected by mutually excitatory connections: e.g. the /p/ unit will be connected to all words that have a /p/ in the current position. These connections make the model very interactive. Not only can units at a lower level activate units at a higher level, but also the other way around: a word unit that receives top-down (i.e. syntactically, semantically or pragmatically determined) activation will in turn activate lower-level representations that are consistent with it, and these will then inhibit representations on their own level. Through its top-down excitatory links, Trace can account for lexical effects on sublexical processing, such as phoneme restoration.

While Trace uses phonemes and features as intermediate units of representation, other units could have been used. McClelland and Elman could have chosen syllable units instead of phoneme units, or even distributed representations (where perceptual objects are represented by patterns of activation across nodes) instead of local ones (where perceptual objects are represented by the processing units themselves). But as with the Cohort model, this inherent openness of Trace with regard to the units of representation has to be given up in any implementation. As input representations to their model McClelland and Elman use featural representations based on a segmental transcriptions.

Conclusions

Word recognition models are often (deliberately) vague about the nature of the representations used, particularly about the nature of the input representations. This is deliberate because the focus has been more on the processing architecture than on representational questions.

Nonetheless, lexical representations are generally assumed to be composed of smaller units, most commonly phonemes. In the Cohort model (and also in the Neighbourhood Activation

model; see Luce et al., 1990, Luce and Lyons, 1998) this is implicit in the way the competitor set of a lexical representation is computed from phonemic transcriptions. In Trace, the lexicon is the network of word-, phoneme- and feature-nodes; consequently lexical representations can be regarded as composed of phonemes and ultimately features in Trace. Shortlist, another recent connectionist model (see Norris, 1994), also assumes that lexical representations are composed of phonemes.[3]

The input to the models are in most cases strings of phonemes. The recent PARSYN model (Luce et al., 2000) is the exception: its input are strings of allophones. The choice of input representations is sometimes solely due to convenience, as in the Cohort model, and sometimes it is theoretically founded, as is the case for PARSYN where allophonic representations are proposed to model putative effects of phonotactic probability. In general, models tend to use input representations that are segmented into the same units that are also used to segment lexical representations; and in most models this unit is the phoneme. This general tendency is made explicit by Norris (1994, p 225): "[W]hatever form the input to the model takes, there must be an explicit form-based lexical representation of words expressed in the same vocabulary. The form-based representation is essential for the working of the model because the competition mechanism depends crucially on being able to align lexical candidates with the input." In other words, the input to the model must be specified in a form that is *commensurate* with the way that lexical representations are specified. I will revisit this requirement in the next chapter.

1.2.2 The bottom half: speech perception models

One of the main goals of models of speech perception is to solve the 'problem of lack of invariance', i.e. to deal with the variability in the way the same phoneme is produced, both across and within speakers. A major point of contention in this field has been whether the objects of speech perception are acoustic or articulatory in nature.

Pisoni and Luce's (1987) account of word recognition described at the beginning of this chapter is an example of an *auditory theory* of speech perception and word recognition. Models such as this one have also been referred to as *information-processing* models (see e.g. Goldinger et al., 1996), because they tend to have several levels where information extracted from the speech signal is processed and transformed. There are many models of this kind: Studdert-Kennedy (1974, 1976), Oden and Massaro (1978), Diehl and Kluender (1989), Nearey (1990), Kluender (1994), Ohala (1996).

Slightly different but also stressing the acoustic/auditory nature of speech perception is Ste-

[3] Unlike Trace, Shortlist is an autonomous and not an interactive model; and it uses a more realistic way of modelling the time course of auditory word recognition.

vens' theory of acoustic landmarks (Stevens and Blumstein, 1978, Blumstein and Stevens, 1979, 1980, Stevens, 1989 and particularly Stevens, 2002). The central idea is that because of the structure of the vocal tract and the way speech is produced, it contains acoustic landmarks (discontinuities, peaks and valleys in the spectral representation) that are sufficient to identify distinctive features and ultimately segments. Stevens and his co-workers claim that, despite the large amount of variance, the speech signal contains enough invariant information to identify abstract units. The acoustic landmark model of speech perception thus fits well into Pisoni and Luce's account of word recognition.

The major alternatives to the auditory models are the Motor Theory (Liberman et al., 1967, Liberman and Mattingly, 1985) and Fowler's Direct Realist model (Fowler, 1986, Fowler and Rosenblum, 1991). Both these models distinguish *proximal* from *distal* objects of perception; they agree with the auditory models that the proximal objects of speech perception are auditory, but the distal objects, they claim, are vocal tract gestures or intended gestures in the case of the revised Motor Theory (Liberman and Mattingly, 1985). To accept gestures as the objects of perception could solve the problem of lack of invariance, if the gestures could be shown to be invariant; gestures can also easily explain the integration of multiple cues into a single percept, as well as trading relations between these cues (even across modalities). Gestures could, for example, explain the famous McGurk-effect, where images of velar plosives combined with the sound of bilabial plosives produces an alveolar percept. This finding is difficult to explain for a purely auditory theory.

While the Motor Theory and the Direct Realist model *could* solve these problems, their proponents have so far failed to explain how the direct perception of distal vocal tract gestures is possible. Even in a spatial domain, where it make sense to say that *we* are perceiving physical objects and not sensory data (the light that is reflected off the object and impinges on our retina, the sounds that emanate from the object, etc), our *perceptual apparatus* arguably still has to recover the object from the sensory data. In addition, it is not obvious that articulatory gestures (or mental states about these gestures) are indeed distal objects of perception in the same sense that, say, apples, people and trees are.

1.2.3 The whole: mediated lexical access

Whether articulatory gestures can be the appropriate objects of perception need not further concern us. What is important for our purpose is that all speech perception models discussed so far assume that the objects of perception are smaller than words (features, segments or articulatory gestures), and that all word recognition models discussed assume that the input is a string of smaller units (segments or features). The two types can thus complement each other

nicely. Speech perception models can provide the input to word recognition models, while solving the problem of lack of invariance. Word recognition models can focus on lexical activation and the competition between lexical representations.

I will refer to this division of labour as *mediated lexical access*, and call the corresponding type of model a *mediated-access* model.[4] How to best define the term *mediated*, is discussed in the next chapter, particularly §2.6.

1.3 An alternative solution: direct lexical access

In this section I will discuss an alternative type of word recognition model where access to the lexicon is direct, i.e. without a stage of processing where smaller units such as features or segments are recognised. I will start with an early model that was developed specifically with auditory word recognition in mind (§1.3.1) and then look at two general memory models that can be used as models of the mental lexicon (§1.3.2). The presentation of these alternative models will be a bit more detailed, mainly because they are less well known than the models discussed so far.

1.3.1 Lexical Access from Spectra

Klatt with his Lexical Access From Spectra (LAFS) model (Klatt, 1979, 1989) proposes that there are no intermediate stages of processing, and that lexical representations are compared directly with the (transformed) speech signal. Klatt assumes that the signal is transformed into a spectral representation. Lexical representations therefore need to have a spectral form too: they are sequences of normalised spectral templates.

In the 1979 version of the model, the whole lexicon is one large network. Paths through this network represent words and utterances. Initially, the nodes in the network correspond to phonemes. These will then be replaced by phonetic nodes with the help of phonological rules; e.g. the word AND is represented by its phonological form /ænd/ which will be replaced by its possible phonetic realisations (e.g. [ænd], [ən] or [n̩]). Finally, each of these nodes will in turn be replaced by a subnetwork of spectral templates. The spectral templates themselves are based on diphones, so that they encode the transitions between adjacent phonetic nodes. An utterance such as 'put it down' is thus represented by the templates for the [*silence*-p] diphone followed by the templates for the [p-ʊ] diphone, followed by the templates for the [ʊ-t] diphone, followed by the [t-ɪ] diphone, etc.

In order to compare lexical representations and input representations, a spectral similarity

[4] McLennan et al. (2003) were, to my knowledge, the first to use *mediated access* in this way.

(or distance) metric is required. The path through the network which is most similar overall will be the word or sentence that is recognised. Note that while Klatt's model uses segments (phonemes, conditioned allophones) for the initial construction of the network of spectral template, these units are not used in the recognition process.

Klatt's model grew out of research on spoken word recognition by machine, where implementational detail is very important. For our present purpose, however, it is the general theoretical principle embodied in LAFS that is of interest. Instead of trying to 'undo' contextual variation at different levels of processing (as suggested by Pisoni and Luce 1987) Klatt suggests compiling the variation into the lexicon; i.e. instead of having one (phonemic) lexical representation per word, Klatt's model has as many representations as there are different phonetic realisations. The problem of lack of invariance is thereby bypassed. Another advantage of this approach is that decisions are not made too early. Since there is no segmental level of processing, wrongly identified segments will not cause word recognition to fail, as they may in a sequential model such as Pisoni and Luce's.

In the models discussed earlier lexical access is *mediated*; in Klatt's model lexical access is *direct*. And while the former solution to the problem of lack of invariance may be called a *processing solution*, Klatt proposes a *storage solution*: variation is not undone previous to lexical access, but is retained and used to enable lexical access.

1.3.2 Multiple-trace models

A more radical interpretation of the storage solution can be found in various forms of multiple-trace (or exemplar) memory models. Multiple-trace models postulate that every experience will leave a trace in memory that resembles the experience; new experiences are identified and categorised by comparing them with all the traces of previous experiences. Several multiple-trace models have been proposed (e.g. Medin and Schaffer, 1978, Feustel et al., 1983, Hintzman, 1984, Nosofsky, 1988, Kirsner et al., 1987, Kruschke, 1992, Estes, 1993). I will only present two: Hintzman's MINERVA 2 model (Hintzman, 1984, 1986, 1988), as an example of a pure exemplar model; and Kirsner et al.'s record-based model of word recognition (Kirsner et al., 1987), as an example of a model where memory traces also retain the products of previous analyses.[5]

Minerva 2

MINERVA 2 is a model of memory. It represents experiences in terms of sets of properties. Memory traces are records of experiences and retain, though not necessarily perfectly, the configuration of properties that constituted the experiences that caused them. An experience first

[5] MINERVA 2 has been applied to speech by Jusczyk (1993), Jusczyk (1997, ch. 8), and Goldinger (1998).

creates a trace in primary or short-term memory, which then sends a probe (or retrieval cue) to secondary or long-term memory. This probe activates all memory traces residing in long-term memory in proportion to their similarity to the probe; the total activation of all traces is sent back as an 'echo' to short-term memory. The intensity and content of this echo determine how the new experience is identified.

When a memory trace is activated by a probe, its activation level will spread to all its properties, even those that are not shared by probe and trace. This has the consequence that the echo that is sent back to short-term memory may contain properties not contained in the probe and thus not in the experience itself. It is in this way that MINERVA 2 can account for associative learning and abstraction: associative learning is explained by echoes that contain different properties than their experiences, and abstraction by echoes which are more generic than the experiences.

Formally, MINERVA 2 works as follows. *Experiences* produce *traces*; these are represented as vectors of the integers $+1$, -1 and 0. Each position in such a vector represents one property, with $+1$ indicating that the property is activated, -1 that it is inhibited, and 0 means that the property is neither activated nor inhibited. The *probe* is represented by the vector P_j, and all the *traces* stored in long-term memory are collectively represented by the matrix T_{ij}, where $i = 1 \ldots n$ stands for the set of all traces and where $j = 1 \ldots m$ refers to the number of properties.

The *similarity* of the i-th trace to the probe is defined as

$$S_i = \frac{1}{N_R} \sum_{j=1}^{N_R} P_j T_{ij}, \tag{1.1}$$

where N_R is the number of properties that are relevant to the present comparison. Each property of the i-th trace is multiplied by the corresponding property of the probe. This yields $+1$ if they are either both $+1$ or both -1, -1 if the property of the probe and the trace have opposite values, and 0 if either trace or probe have the value 0 for that property. Summing over all properties yields the total activation level of trace T_i, and dividing by N_R normalises this value.

The *activation* of the i-th trace by the probe is defined as

$$A_i = S_i^3. \tag{1.2}$$

The power function is required to make traces that are similar to the probe have a much higher activation level than traces that are less similar to the probe, thereby increasing the signal-to-

noise ratio of the resulting echo.

The echo has both an intensity I and a content C_j. *Echo intensity* is given by

$$I = \sum_{i=1}^{m} A_i; \tag{1.3}$$

i.e. the intensity of the echo is equal to the total activation level of all traces. And *echo content* is given by the vector

$$C_j = \sum_{i=1}^{m} A_i T_{ij}. \tag{1.4}$$

Notice that echo content is computed by property: the echo content of the j-th property is defined as the sum of the activation of all instances of that property in T_{ij}; hence echo content is a vector with one value per property.

Since echo intensity is determined by the total amount of activation in long-term memory, it can be regarded as a measure of the *familiarity* of the probe: more familiar experiences will produce a higher echo intensity than less familiar or new experiences. Echo content, on the other hand, is a measure of the *specificity* of the echo. If only a small number of traces that share much of their properties with the probe are activated, the echo content will be specific and similar to the content of the probe. If a large number of traces are activated – including many that only share some properties with the probe vector – the echo content will be much more diffuse and dominated by those properties that are shared by most traces, resulting in a more generic or abstract echo.

MINERVA 2 is a pure exemplar model, as it does not propose any abstraction processes at the time of storage, such as the formation of category prototypes. One of the main purposes of MINERVA 2 was the demonstration that effects that are normally interpreted as evidence for abstract mental categories can be accounted for entirely on the basis of non-abstract representations (Hintzman, 1986).

Kirsner et al.'s record-based model

Like MINERVA 2, this record-based model of word recognition (Kirsner et al., 1987) assumes that every experience produces a memory trace and that new experiences are categorised by comparing the trace produced by the experience to all traces in long-term memory. Where the two accounts differ, however, is in their assumptions regarding what is stored in a memory trace. Whereas in MINERVA 2, traces consist only of experiential properties, Kirsner et al. propose that memory traces (or *records*, as they call them) may also contain information which is

more abstract, such as a phonemic analysis of the input.

In the record-based model, experiences produce *descriptions*. Descriptions can be compared to files: they are places where information about something – in the present case experiences – are stored. Descriptions stored in long-term memory, are called *records*. The content of descriptions and records are *codes*. Codes are similar to MINERVA 2's properties, except that they are not restricted to experiential properties. In the record-based model codes may be produced at different levels of analysis. A record of a word that has been perceived in the past will, at least, contain codes referring to its sensory shape (auditory in the case of spoken words and visual in the case of written words); but in addition to this sensory information it may also contain phonemic, graphemic or morphemic codes, depending on the kind of analyses that have been performed when the experience was first processed.

When a new experience is made, the description it produces will be compared to the records stored in memory, and its categorisation will depend on its similarity to these records. Because records contain (or may contain) information gathered on many different levels of analysis, the categorisation of new experiences will not only depend on the experience itself but also on kinds of analyses that are carried out.

Despite the numerous differences between Hintzman (1986) and Kirsner et al. (1987), there are three properties that they both share and that, therefore, may be regarded as defining characteristics of multiple-trace or exemplar models of memory:

1) Categories are represented in memory not by one single representation but by a collection of representations.

2) These representations are copies or traces of previous experiences.

3) Representations are combined and experiences categorised on the basis of similarity between the traces.

MINERVA 2 is the more radical multiple-trace model of the two, because it does not use any process of abstraction. Effects that look like they are due to abstraction are explained as experiences that produce generic echoes. This is the main claim embodied in MINERVA 2: that a model without active abstraction can nonetheless account for most, if not all, seemingly abstract behaviour. What makes MINERVA 2 very attractive as a model of word recognition is that it allows us to determine how far a pure multiple-trace model can take us in explaining perceptual and cognitive phenomena that seemingly involve abstraction.

Several models based on MINERVA 2 have been proposed in speech perception research (Jusczyk, 1993, 1997, Goldinger, 1998). In all of these implementations, words are stored in

long-term memory as traces instead of as abstract representations. Johnson (1997) uses a similar model (Nosofsky, 1988) to account for the perception of vowels. His implementation proposes that vowels are stored as collections of exemplars and not as abstract or prototypical phonemic representations. This makes it possible that exemplars of different sizes could be used in parallel, corresponding to the processing stages of a serial model; for instance a lexical level where words are stored as exemplars, and a segmental level where exemplars correspond to phonemes.

Multiple-trace models may seem rather implausible: surely we cannot store all our experiences in all their detail? This alleged implausibility is alleviated by the fact that multiple-trace models also take account of forgetting. In MINERVA 2, for example, stored traces lose strength over time and their properties eventually revert to zero. While every experience indeed leaves back a trace, we do not need an infinite amount of storage space, thanks to the process of forgetting.

In addition, experiments which presented subjects with stimuli from different speakers have shown that details about a speaker's voice are retained over a long time (e.g. Craik and Kirsner, 1974, Schacter and Church, 1992, Palmeri et al., 1993, Church and Schacter, 1994, Goldinger, 1996). In the task used by Craik and Kirsner (1974) subjects had to classify words as *new* (i.e. not presented before) or *old* (i.e. presented before). If first presentations and repetitions were by the same speaker, repetitions were more accurately identified as *old* than when they were presented by different speakers. This basic finding has been replicated with different kinds of tasks and many more speakers; it has also been shown to be fairly long-lasting (e.g. Goldinger, 1996). These effects of speaker voice do not prove that access to the lexicon has to be direct and not mediated, but they at least make a direct account where whole words are stored in the lexicon as fine-grained auditory representations plausible.

Finally as e.g. the studies by Hintzman (1986, 1988) and Nosofsky (1988) have demonstrated, multiple-trace models can account for many effects which appear to involve abstraction. Multiple-trace models are therefore not only plausible, but they are also in principle capable of explaining abstractive and symbolic behaviour.

1.4 Conclusions

In the preceding pages, I have discussed two radically different ways of approaching auditory word recognition. Within these two broad types there is considerable variation between models, and word recognition models can be classified according to other criteria not discussed in this chapter (e.g. whether they are interactive or autonomous). Two aims have guided this

discussion. The first was to present a survey of word recognition and speech perception models that focuses on representational issues. The second was to demonstrate that the question my thesis tries to address, namely whether auditory word recognition is direct or indirect, is relevant to existing models of auditory word recognition. The two types of models we have found are best described as *mediated-* and *direct-access* models.

Mediated-access models propose that auditory word recognition is a process that involves several stages of processing. This has been explicitly formulated by Studdert-Kennedy (1974, 1976) and Pisoni and Luce (1987), among others; and it is presupposed by the standard division of labour between speech perception models (which mainly address the problem of lack of acoustic invariance) and word recognition model (which assume that problem to be solved, and focus on issues of lexical activation and competition). *Mediated-access* models attempt what I have called a *processing solution* to the problem of lack of invariance, and to speech perception and word recognition in general.

Direct-access models have no stages of processing intervening between the input and the lexicon; instead they propose that lexical representations have a form which makes them directly comparable to input representations. This is what I have called a *storage solution* to the problem of lack of invariance. Instead of having an intermediate stage of processing that produces units that provide invariance (such as phonemes or features), the storage solution solves the problem of lack of invariance by retaining the variance in the lexicon, either by compiling it into the lexical representations (Klatt's LAFS model) or by storing traces of all previous occurrences (multiple trace models such as MINERVA 2).

2 / Descriptive dimensions

The previous chapter presented a short overview of models of auditory word recognition with particular attention to representational question. The present chapter follows on from this and takes a more systematic look at representational issues in auditory word recognition. The ultimate goal is an inventory of the theoretical options available in the form of a typology of auditory word recognition models (see Chapter 3). To get there, we need to identify the relevant dimensions along which word recognition models and the representations they use can be classified.

To facilitate the discussion, I will first introduce a *generic model* of spoken word recognition with a prelexical and lexical level of processing. This model is not set up in competition to the models described in Chapter 1, but only as a tool to highlight representational issues. The model therefore focuses on the representations used and not the processes involved in word recognition.

2.1 A generic model of auditory word recognition

The generic model represented in Figure 2.1 has four stages or levels of processing: auditory, prelexical, lexical and postlexical, each having some form of representation of the acoustic signal as its input and another, transformed, representation as its output. It is somewhat reminiscent of the linguistic model of Pisoni and Luce (1987), except that the generic model in Figure 2.1 has only one (segmental) prelexical level, while Pisoni and Luce's model has two: one featural and one segmental. This difference is not crucial for the following discussion.

At the *auditory stage* the incoming acoustic signal is transformed into an auditory representation, determined by the properties of the human hearing apparatus. At the *prelexical level* segmental representations (phonemes or phones) are extracted or recognised. At the *lexical level*, these segmental representations are assembled into lexical representations. Words that have been recognised are then passed on to the higher (i.e. syntactic, semantic and pragmatic)

processing modules; this is the *postlexical stage*.

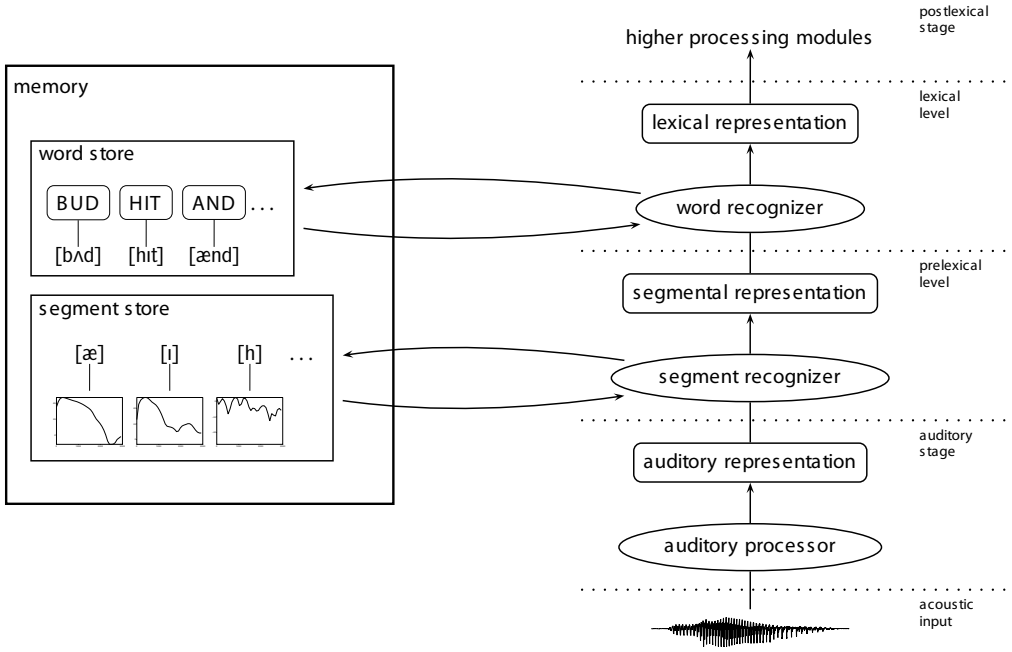


FIGURE 2.1: The generic model of auditory word recognition. For an explanation of the difference between processing stages and processing levels see the running text.

The model distinguishes processing *stages* from *levels* of representation. This difference will be explained further in §2.6. For the present the following description should suffice. The term processing stage is the broader; it refers to whatever can be regarded as a distinct step of processing. The auditory processing stage, for example encompasses all transformations that take place in the auditory system.

The prelexical and lexical levels are also processing stages; but in addition they are stages at which units of a particular size and abstractness are *being recognised*. At the prelexical level, these units are segments, but could also be smaller units such as features or larger units such as syllables; and at the lexical level the units are words. Because stretches of the speech signal need to be recognised as being one type of unit and not another – e.g. a /b/ as opposed to a /p/ or an /f/, or the word *cat* instead of *cap* – we require a memory store for each of these stages of processing, where the units available for recognition reside. What distinguishes a level from a stage is thus that a recognition process takes place, and the existence of a corresponding store.

FIGURE 2.2 shows in schematic form how the generic model should be understood to process an incoming speech signal. The first step is the *auditory* processing of the acoustic signal. We may assume that the incoming signal is transformed into a continuously updated auditory

spectrum. Readers may substitute their preferred auditory representations and transformations here: the important thing is that at the auditory stage, no units of any size are extracted from the signal; the signal is merely transformed according to the constraints of the human auditory system.

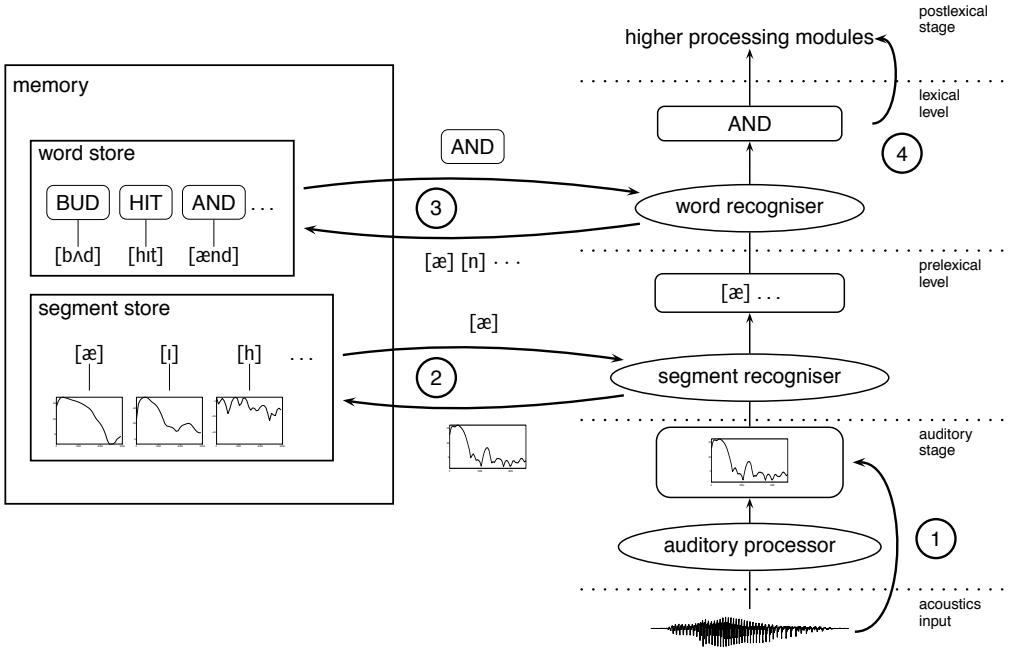


FIGURE 2.2: Processing steps: (1) auditory transformation; (2) segment recognition; (3) word recognition; (4) higher-level processing

The extraction of units takes place at the next stage, the *prelexical* level of processing. In the diagram, the units that are extracted or recognised are segments; these may be phonemes or phones. Recognition takes place by comparing the continuously updated auditory spectrum against spectral templates stored in a *segment store*, one for each segment. Again, any other suitable process may be substituted for this. The output of this operation is a sequence of abstract segmental representations. We can extend the prelexical level to incorporate units both smaller or larger than the segment. For example, we could split it into two levels: a level of feature recognition followed by a level of phoneme recognition. What is important is that at the prelexical level a continuous signal is transformed into a string of sublexical units. These units need not be linguistic units, but in most models they are; in our illustration, they are segments.

The third step is the recognition of words from sequences of segments. The word recogniser module takes the string of segments that the phoneme recogniser produces and compares them to the lexical items stored in the *word store* or lexicon. The stored lexical items can be conceived

as pairs of phoneme strings and lexical representations, where a lexical representation would consist of all the information that is required to further process the recognised word (i.e. syntactic, semantic and pragmatic information).

Once a word has been recognised, the information associated with it is communicated to the higher processing modules that constitute the postlexical processing stage. This is the last of the four steps identified in FIGURE 2.2, and will not be dealt with here.

I want to reiterate that the purpose of the generic model just described is solely to facilitate the discussion of the descriptive dimensions; it is not meant to embody any theoretical claims. The most important descriptive dimensions will be whether the model has a prelexical level or not (see §2.6); this means that the prelexical level in FIGURE 2.1 and FIGURE 2.2 is a point of contention and should not be understood as an essential ingredient of the generic model.

2.2 Descriptive dimensions: overview

When describing anything, it is obviously important to choose appropriate descriptive terms, as they determine both the substance and the limits of the description. Choosing inappropriate terms may have severe consequences; in the worst case it can be an obstacle to an adequate description. There are no hard and fast rules about choosing descriptive terms; it is only when we try to apply them that we can see whether our choices have been fruitful.

The descriptive terms and dimensions used in this thesis are based on the following four questions (repeated from the introduction):

1) Are lexical representations unstructured wholes, or are they composed of smaller, sublexical representations, such as syllables, phonemes or distinctive features?

2) Does the speech signal map directly onto these lexical representations, or are there prelexical levels of processing?

3) If a prelexical level of processing exist, do the representations used at the prelexical level correspond to the sublexical representations used in the lexicon?

4) How abstract are the representations used at any given level of processing?

In our discussion of multiple-trace models we came across another relevant issue, which shall be our fifth question:

5) Do representations at any given level consist of a single summary item, or are they collections of individuals (traces or exemplars)?

Based on these questions, I propose three *representational* dimensions (i–iii) and one *architectural dimension* (iv):

i) **abstract/concrete:** the degree to which a representation is abstract or concrete;

ii) **exemplar/summary:** the degree to which a representation is a collection of exemplars or a summary representation;

iii) **structured/unstructured:** whether a representation is structured into smaller units or remains unstructured;

iv) **prelexical levels:** whether prelexical levels are needed for auditory word recognition.

Note that the term *dimension* is used in rather loose way, as only the first two are continuous. The fourth, architectural, dimension is essentially a yes-no question: is there a prelexical level or isn't there? For the third, several layers of structuring are possible. The first three are called *representational* dimensions because they describe representations and can be applied to any type of representation regardless of size or stage of processing they are used at. The *architectural* dimension refers not to representations but to the structure of whole models; that is why it is called architectural.

I will first discuss the three representational dimensions. To avoid confusion, it is important to realise that the term *representation* can refer to two things (see the generic model in FIGURES 2.1 and 2.2): first, the input or output of a processing stage (e.g. the auditory spectrum that is the product of the auditory stage, or the segments that is the output of the prelexical stage); and secondly, what is stored in memory. The later type of representation consists of pairs of the former, i.e. an auditory spectral template and its corresponding segment, or a sequence of segments and their corresponding lexical representation. I will generally use the term representation in the first sense, because it is the more widely applicable.

2.3 The abstract/concrete dimension

Models of word recognition may differ with regard to the abstractness of their representations, particularly their lexical representations. Most models use relatively abstract lexical representations, mainly strings of phonemes. Klatt's LAFS model, on the other hand, has lexical representations that are very close to the auditory representations of the acoustic input: spectral templates. We need to define abstractness in a way that covers these cases and corresponds to, but also elaborates upon, the common usage of the term.

The core meaning of *abstract* is 'distant from physical reality'. When we make an abstraction, a description (or in our case representation) is made less complex by omitting detail – detail

which can be considered superfluous and irrelevant for the kind of description needed. The core meaning of abstraction is thus a reduction in complexity. The complexity of a representation is defined by:

a) the number of independent, i.e. unrelated, variables used to define a representation: the more variables used, the less abstract it is;

b) the number of independent variables in the whole domain (i.e. the whole language); this number may be the same or higher than that in clause *a*, but never lower;

c) the number of distinct values that the variables in clause *a* and *b* can take.

That this definition captures the usual meaning of *abstract* can be seen with the help of the examples listed FIGURE 2.3, which shows four levels of abstractness, with row 1 being the most and row 4 the least abstract. Let us start with phonemes and phones (the phonetic realisation of phonemes) first – as shown in row 2 and row 3 – as they are the most familiar.

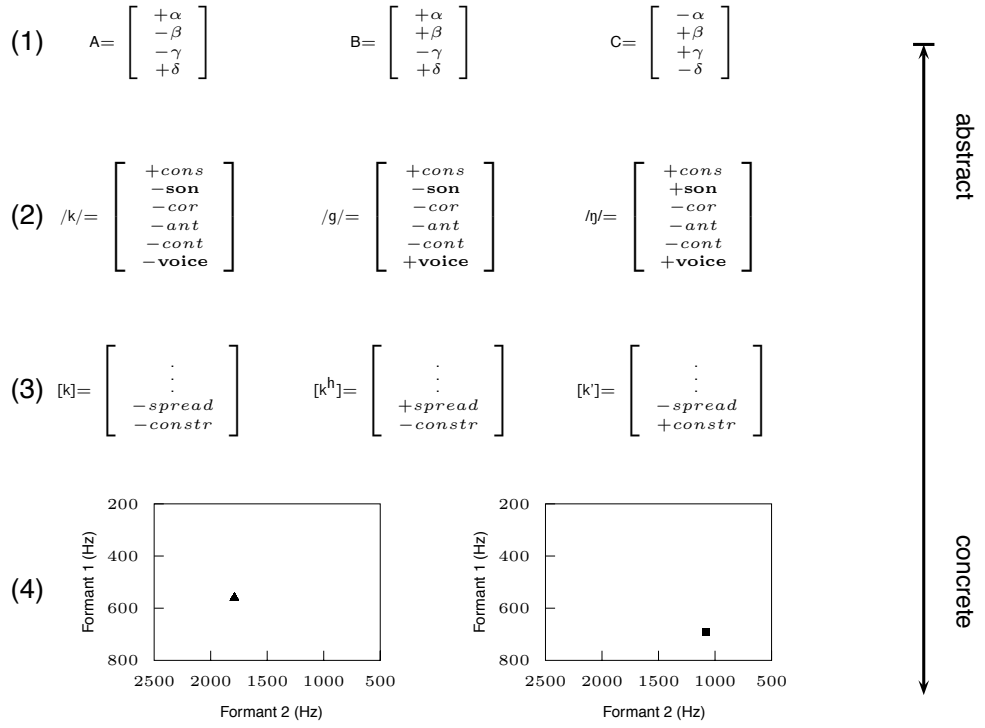


FIGURE 2.3: Examples of the abstract/concrete dimension. Representations get less abstract as we move down from row 1 to row 4. Rows 2 and 3 show phonemes and (allo-)phones, respectively. Row 1 gives the most abstract segmental representations possible, and row 4 shows representations which are more concrete than phones.

Let us assume a hypothetical language with only three consonantal phonemes, /k/, /g/, and /ŋ/, all listed in row 2. The phonemes are described using the six features $\pm cons$ (consonantal), $\pm son$ (sonorant), $\pm cor$ (coronal), $\pm ant$ (anterior), $\pm cont$ (continuant) and $\pm voice$ (voicing).[1] Let us further assume that the voiceless stop /k/ has got the three allophones listed in row 3: [k] (unaspirated), [k$^{\text{h}}$] (aspirated), and [k'] (ejective). To describe these allophones or phones[2] we need two additional features; let us call them $\pm spread$ (for spread glottis) and $\pm constr$ (for constricted glottis).[3] This is an illustration of clause *a* of the definition. To describe /k/ we need two features less than to describe e.g. [k$^{\text{h}}$]: therefore, /k/ is more abstract than [k$^{\text{h}}$]. This is as we want it to be.

An phone can never be described using less features than its *corresponding* phoneme: to describe [k$^{\text{h}}$] we need at least as many feature as we need to describe /k/; but this is not true for any *randomly selected pair* of phones and phonemes. Let us assume that our hypothetical language only has one vowel phoneme /a/, with the two allophones [a] and [e].[4] To distinguish the phoneme /a/ from the consonants, we only need one feature, namely $\pm cons$. To describe the two phones we need two features: $\pm cons$, and $\pm hi$ (high) to distinguish the two phones from each other. This means that we need only two features to describe the phone [a], but three features to describe the phoneme /k/. Clause *a* of our definition would imply that the phone is more abstract than the phoneme. This is not what anyone would want to claim, however; what we want to say is that *all* phonemes are more abstract than *all* phones. This is what clause *b* is needed for. The overall numbers of features needed to capture all phones in a language cannot be lower than the number of features required to capture all phonemes, because no language can have fewer phones than phonemes as each phoneme has by definition at least one phone.

To understand why clause *c* of the definition is needed, we have to compare the featural representations of phonemes and phones with the formant-value representations shown in row 4 of Figure 2.3. Let us assume that like the representations in the previous rows, these also represent segments, but using only two variables: first formant (F_1) and second formant (F_2). By

[1] The features used here go back to Chomsky and Halle (1968), and have since become known as the SPE feature system. Note that in this hypothetical case the features $\pm cor$, $\pm ant$ and $\pm cont$ are redundant; only the three features $\pm cons$, $\pm son$ and $\pm voice$ are required for a unique description of all the consonant phonemes of the language.

[2] I use the terms allophones and phones in the following way: *phone* refers to the phonetic realisation of a phoneme; *allophone* refers to the special case where a phoneme has two or more noticeably different phonetic realisations which may be conditioned (such as clear and dark /l/ in most varieties of English) or in free variation (such as the various ways in which syllable-final /t/ may be produced in British English). See e.g. Trask (1996) or Clark and Yallop (1995, pp. 91–99) for brief introductions.

[3] These terms were introduced by Halle and Stevens (1971), but have been appropriated into the SPE feature system. Chomsky and Halle (1968) used the terms *heightened subglottal pressure* and *ejection* instead.

[4] No such language exists. The minimum number of vowels seems to be three; smaller numbers have been reported but also disputed. See e.g. Maddieson (1984, p. 126).

clause *a* these representations are less complex, and thus more abstract, than the featural representations in rows 2 and 3, which use six and eight variables respectively – or three and five if we only count non-redundant features. Most speech scientists would probably want to regard an F_1/F_2-representation as more concrete than a representation in terms of phonetic features, because an F_1/F_2-plot allows the very precise location of vowels in an acoustic space. Clause *c* of the definition (the numbers of values that a variable can have) takes care of this. F_1/F_2-values can be regarded as continuous, while distinctive features are binary.

I hope to have shown that the definition of *abstractness* in terms of complexity captures and elaborates upon the common meaning this term has in phonetics and phonology. Two points remain. The first is that the definition has three parts, and it is a fair question how the parts combine and which takes priority. This is partly a practical issue. If we want to distinguish phonemes from phones, then clause *b* is the most relevant (as we have just seen). If we wanted to determine which of two allophonic representations should be regarded as the more abstract, then clause *a* is most relevant. And if we want to compare different acoustic representations, we need a combination of clause *a* and clause *c*.

The second point is that our definition of abstractness is continuous, going from the fully concrete – the acoustic signal itself – to the fully abstract – a purely symbolic representation. Such a purely symbolic representation is shown at the very top of Figure 2.3. The only difference between this and the phonemic representation in row 2 is that phonemic, or distinctive, features tend to still have a phonetic interpretation, i.e. they can be related to properties of the speech signal. The features used in row 1, on the other hand, are meant to be devoid of any phonetic content. Their purpose is to distinguish and group the units on this level of representation: we can say that A is distinct from B because A is $-\beta$ while B is $+\beta$, and that A has more in common with B than with C, etc. This purely formal definition would be the highest possible level of abstraction, as it would only use as many features as are required to distinguish different representations.

2.4 The exemplar/summary dimension

The exemplar/summary dimension describes representations in terms of whether they are single objects or collections of objects. A vowel phoneme such as /i/ can be represented by prototypical (or mean) F1- and F2-values, as illustrated on the left side of Figure 2.4, row 2; this would be a *summary representation* of that phoneme. Alternatively, /i/ could be represented more directly by the F1- and F2-values of all instances of /i/ heard so far, as illustrated on the right side of row 2; this would be an *exemplar representation* of /i/.

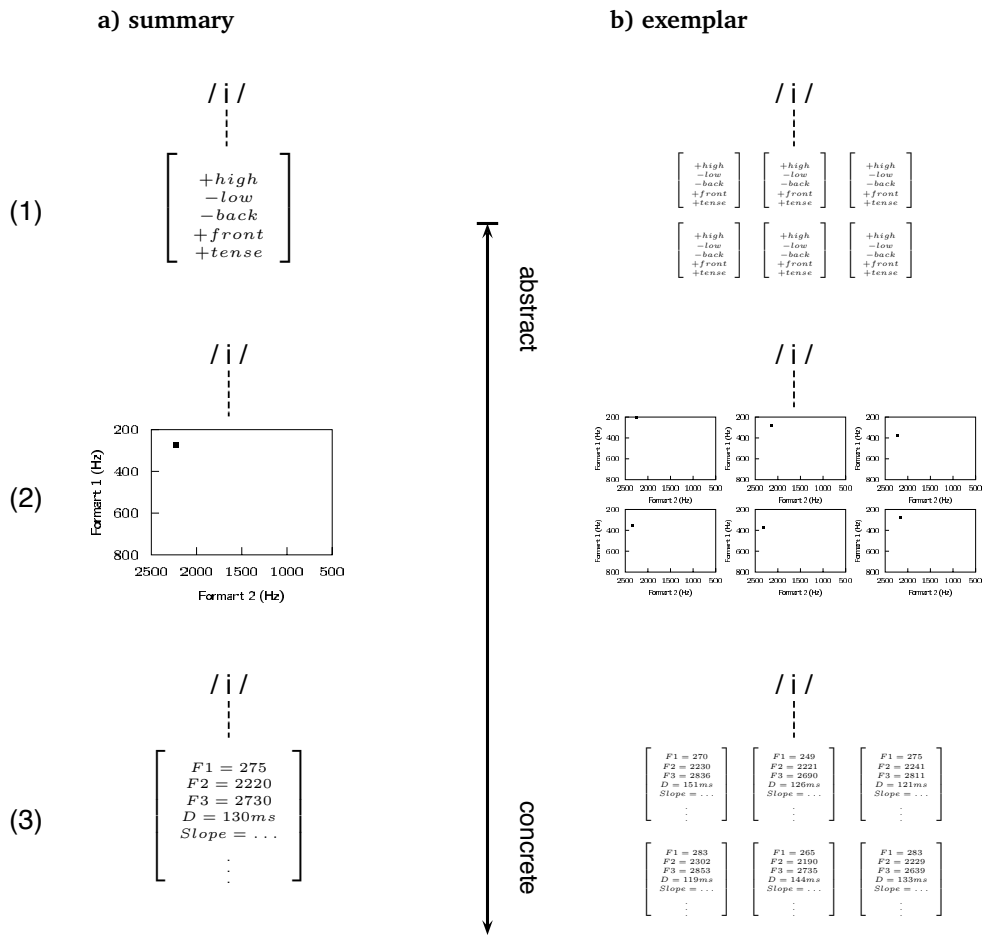


FIGURE 2.4: Examples of the exemplar/summary dimension at different levels of abstractness (see running text for discussion).

The exemplar/summary dimension is more difficult to define than the abstract/concrete dimension. It is not clear whether it is a binary or continuous dimension. There is a sense in which *exemplar*, as used here, is a gradient term: a representation consisting of two million tokens is more exemplar than one consisting of only twenty. On the other hand the term *summary* calls for a categorical definition: either a representation is a sum or average of tokens, or it is not.[5]

A simple way of dealing with this problem would be to regard representations that assign two or more tokens to each type as exemplar representations, and representations with a one-to-one relationship between tokens and types as summary representations. This does, however, not capture the essence of the exemplar/summary difference. A better definition would be one

[5] It would be possible to have exemplar representations whose exemplars are themselves summary representations; e.g. when we take averages of ten F1-/F2-values and then store these individually. This does not make the term *summary* any less categorical, though.

that regards exemplar representations as essentially open; even though there will be physical limitations on storage, what makes a representation a genuine exemplar representation is that it is always possible to add another exemplar to the stack, and this addition will make a difference to the representation. The more exemplars the representation contains, the smaller the influence of a new token will be, but each individual token matters in the overall computation of the type.[6] A summary representation, on the other hand, is fixed: nothing can be added, and the representation can only be changed by replacing it with a new summary representation. In short, exemplar models are *asymptotic* and summary models *categorical*.

In the literature, exemplar models are often regarded as the opposite of abstractionist models (explicitly by e.g. Tenpenny, 1995, Pallier et al., 2001, Pierrehumbert, 2003). There are reasons for this, as we will soon see. But conceptually the two dimensions are clearly distinct, and exemplar representations are *not* the opposite of abstract representations.

This is obvious if we consider that exemplar and summary representations may occur at different levels of abstractness. The vowel /i/ may be represented in a summary fashion not by the mean values of F1 and F2 (as in row 2 of FIGURE 2.4), but by a matrix of many more acoustic variables, such as formants, duration, the slope of formant trajectories, intensity, pitch, etc. (see row 3 of the same figure). Such a summary representation would clearly be more concrete, i.e. closer to the auditory level, than collections of F1/F2-exemplars (as shown on the right side of row 2). In short, for every summary representation there exists in principle a corresponding exemplar representation with the same degree of abstractness.

Even though the exemplar/summary and abstract/concrete dimension are conceptually distinct, the two dimensions are not completely independent. While it is possible to keep exemplars of fully abstract representations, the more abstract the representations are the less variable they will be, and the less sense it makes to store them individually as exemplars. To represent a phoneme by hundreds or even thousands of exemplars of its featural description is quite pointless, as these exemplars will all be identical. This case is illustrated in FIGURE 2.4 on the right side of the top row. In most models of word recognition, summary representations are also fairly abstract representations – e.g. phonemes or phones – and this explains why *exemplar* is often juxtaposed with *abstract*.

The exemplar/summary dimension is a dimension that, unlike the abstract/concrete dimension, only applies to the pairwise representations as stored in memory, and not to the outcome of a processing steps. Any representation of the signal at a given moment in time, and at whatever level of representation, will always be a token or exemplar. The exemplar/summary dimension only applies to mappings of representations in memory (e.g. spectral templates and

[6] The reader is referred back to the discussion of MINERVA 2 in §1.3.2.

phonemes as in FIGURE 2.4, row 2); and it distinguishes one-to-one from one-to-many mappings or, as I have argued, open from closed representations.

2.5 The structured/unstructured dimension

The structured/unstructured dimension describes whether a representation consists of smaller building blocks, or in other words whether a representation is composed of sub-representations. Words, for instance, may be represented either as unstructured wholes or as strings of smaller units, such as syllables or phonemes. I will refer to these sub-word units as sublexical representations.

The structured/unstructured dimension is a binary one: a representation either consists of smaller units or it does not. What can vary, however, is

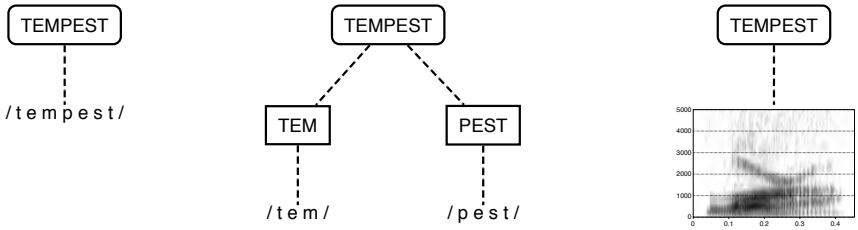
a) the size of these units;

b) whether there is only one level of sublexical analysis (e.g. words divided into segments) or more than one (e.g. words divided into syllables, and syllables into segments);

c) whether the analysis is compositional in the sense that units on one level are fully structured into units of the next-lower level;

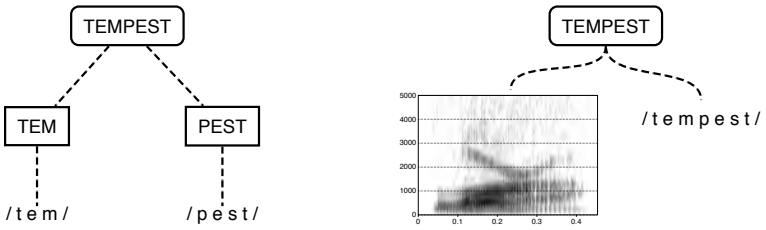
d) whether units on a single level are allowed to overlap or not.

The following display illustrates point *a* (size of the representation) and point *b* (levels of analysis):



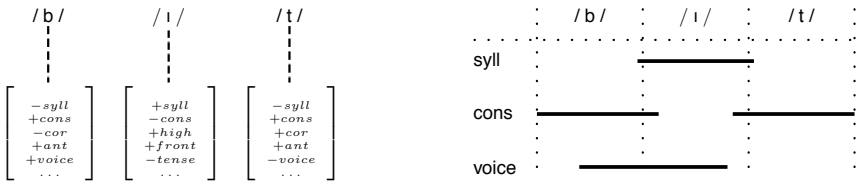
On the left hand side, we have a lexical representation that has one level of analysis, and the size of the sublexical units is that of the segment (phonemes in this case). In the middle, we find a lexical representation with two levels of analysis: syllables and segments. An unstructured lexical representation is illustrated on the far right: the lexical item as a whole is represented in memory as one single whole (a spectrogram in the example).

Compositionality (point *c*) can be illustrated with the following examples:



The left-hand figure shows a compositional representation: the word is fully composed into syllables and these are in turn fully composed into phonemes. The representation on the right is not compositional. The word both has a spectral representation and is divided into segmental subunits; but the spectral representation is not itself composed of segments, or segment-sized spectra. The two representations are thus not on different levels of analysis – though they differ in their abstractness. In a model of word recognition, this would allow for parallel lexical access: one route via phonemes, and the other directly from input to lexicon.

The following display contrasts a non-overlapping analysis on the left, where phonemes are structured into atemporal features, with an overlapping analysis on the right, where the features have a temporal duration and do overlap (as do e.g. the gestures used in Articulatory Phonology; see Browman and Goldstein, 1992):



This is point *d*. Linguistic representations have traditionally been compositional and non-overlapping: each level of representation is assumed to be exhaustively analysable in terms of its next-lower level.[7]

There are many different causes why lexical representations might be structured into sub-lexical representations. Some of them have to do with the internal organisation of the lexicon: the recognition of words may involve a search through a very large store of lexical items; this search could be made much easier if the items were structured. Then, the higher-level processing components may also require lexical items to be structured, e.g. into morphemes. Finally,

[7] Note that in recent non-linear phonological analyses, such as Autosegmental Phonology, this assumption has been relaxed. An example of a representation that cannot be used for exhaustive analysis is the *mora* as a unit of *syllable weight*. Syllables have a certain number of morae depending on the structure of their rhyme, but cannot be parsed into morae.

if auditory word recognition centrally involves a prelexical level where sublexical representations are recognised, it seems inevitable that lexical entries are represented as strings of the same type of sublexical representations: if a sequence of sublexical representations have been recognised, they can then serve to select candidate words from the lexicon.

This means that we can infer from the existence of sublexical representations as units of recognition the existence of the same sublexical representations as units of lexical structure. The converse, however, is not the case. Because there are other reasons – for example the internal organisation of the lexicon, or speech production – why lexical entries may be structured into smaller units such as morphemes, syllables or phonemes, we cannot infer from the existence of subunits in the lexicon that the same type of units will also be used as prelexical representations in the recognition process.

2.6 Levels of processing

The generic model (see FIGURE 2.1 again) is made up of both *stages* (auditory, prelexical) and *levels* of processing (prelexical, lexical). In this section I will try to define what distinguishes a level of processing from a stage of processing, and I will introduce the difference between *direct-* and *mediated-access* models.

I use *stage of processing* as the more general term. A stage of processing can be any of the steps that take us from the acoustic input to the comprehension of what has been said: any processing that is done to the original acoustic input signal. By a *level of processing*, on the other hand, I refer to a stage of processing where units of a particular temporal extension and degree of abstractness are recognised; furthermore, for such a recognition process to be possible a memory store is required. There are two reasons why levels of processing in the sense just introduced are required in auditory word recognition: the first is the *recovery of meaning*, and the second *discreteness*. Let us look at these issues in a bit more detail.

We are relatively free in what we want to call a *stage of processing*. The generic mode of FIGURE 2.1 has an auditory stage of processing, but this stage could be split into sub-stages. It is common to distinguish a peripheral from a central auditory stage (Pisoni and Luce 1987, see also Greenberg 1996). Within both of these we could go further and identify smaller stages still, from the moment when a speech wave impacts on the eardrum until some representation of it – in the form of patterns of discharging neurons – reaches the auditory cortex. Similarly, the postlexical stage of processing covers many distinct processes: syntactic, semantic and pragmatic processing.

What I call *levels of processing* cannot be subdivided in this way. We can think of them as

breaks or jumps in the processing of the speech signal. Such breaks are needed if the goal of speech perception is – as is universally assumed – the recognition of units of meanings, i.e. words. Words as the main units of meanings have two essential properties: (i) they connect acoustic forms to meanings, and (ii) they are discrete.

Meaning cannot be recovered from the acoustic signal simply by transforming it, no matter how many transformations are applied to the signal. To extract the meaning of an utterance, a *recognition process* is required. And this recognition process depends on a *memory store* where forms are paired up with meanings. This can be taken as the basis for defining a level of processing:

> A *level of processing* is a stage of processing where something is being recognised as a unit of representation, where a token is identified as belonging to a stored type.

At least one level of processing is required for auditory word recognition: one where auditory forms are related to meanings (and associated syntactic and pragmatic properties). The store where these form-meaning pairs are kept is the *mental lexicon*. Whether a prelexical level of processing, as depicted in Figure 2.1, is also required is debatable. This level is different from the lexical level, as it does not involve the pairing of forms and meanings. It is a level of processing, nevertheless, because it involves an instantaneous jump from a continuous acoustic representation to a discrete representation. This step also involves a recognition process that requires a corresponding store.

In its simplest form the question whether prelexical levels of processing are required thus reduces to the following. Does the conversion from continuous to discrete take place in the same step as the conversion from form to meaning, namely in the lexicon? Or does the continuous-to-discrete conversion take place previous to the form-to-meaning conversion, on a prelexical level? In principle, there could be many such prelexical levels of processing (e.g. features, phones, phonemes, syllables, etc.), but obviously only one level is required to make the jump from a continuous representation to a discrete representations. I will call models which postulate that a prelexical level of processing *is necessary* for auditory word recognition *mediated-access* models, and those which have no place for a prelexical level of processing *direct-access* models (see Figure 2.5).

Now that the term *level of processing* has been defined and the difference between *direct* and *mediated* lexical access introduced, a few additional remarks are in order. First, it is important to realise that the way I use the term *prelexical* does not imply that prelexical processing has to occur *earlier* than lexical processing; it only has to be *logically prior*. A level of processing is logically prior to the lexical level if lexical processing, i.e. the recognition of words, *requires* the

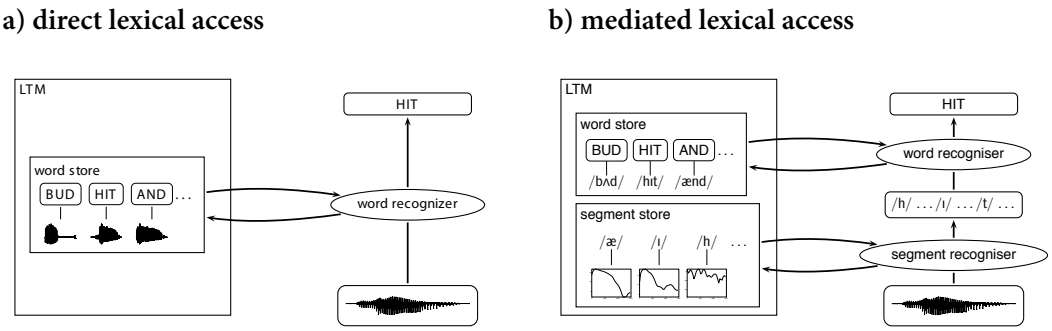


FIGURE 2.5: *Direct-* and *mediated-access* models. LTM stands for long-term memory. Based on the generic model in FIGURE 2.1.

involvement of this level of processing. In addition, the units of processing at this level have to be smaller than words. In sequential models – such as the one by Pisoni and Luce (1987) discussed in the last chapter and the generic model illustrated in FIGURE 2.1 – logical priority implies temporal priority: processing at the prelexical level will occur previous to processing at the lexical level. But this is not the case with interactive models such as Trace (see§1.2.1). This is why a logical definition of *prelexical* is preferable to a temporal definition: because it is the more inclusive.

A second question is what kinds of units count as *prelexical* or *mediating units*. This is partly an empirical question, and the short answer is 'any unit that serves the purpose of mediating between the acoustic signal and the lexicon.' What, then, is the purpose of the prelexical level of processing? One purpose I have already mentioned: the prelexical level can serve to transforms continuous representations into discrete representations. Discrete representations have the advantage of being more manageable than continuous representations, particularly if they can be combined to form larger representations. If we allow phonemic prelexical representations, instead of having to have a continuous-to-discrete mapping for each lexical entry, one mapping for each of the 40-odd phonemes of the English language will suffice.

Another potential advantage of a prelexical level is that it may provide a solution to the problem of lack of invariance, the fact that the same phoneme is produced differently in different contexts. This variation needs to be 'undone' in the recognition process – at least this is the assumption of *mediated-access* models (see §1.2). The prelexical level might be a better place to undo this variation than the lexicon: because the variation is caused by the way speech is produced, it should be constrained by phonetic or prosodic units (such as syllables, segments, features or gestures) and not necessarily words qua units of meaning.[8]

[8]In the word production literature words or lemmas are obviously used as planning units (though Wheeldon and Lahiri (1997) suggest that these units should be *prosodic* words), but it is generally recognised that they

Given these two purposes, especially the problem of lack of invariance, it is not surprising that most *mediated-access* models have chosen some well-established linguistic units as their prelexical unit of processing, the most popular arguably being the phoneme (see my review of word recognition models in §1.2). The phoneme will also be the prelexical unit that my experimental research focuses on (see Chapter 5). But I should add that prelexical units do not necessarily have to be well-established linguistic units; they simply need to be a units which are potentially capable of solving the problem of lack of invariance. This requirement puts a certain limit on the size and abstractness of the units we can reasonably consider. Given what we know about speech production and perception it seems unlikely, for example, that an arbitrary prelexical unit of 1 ms duration could serve this purpose: it is simply too short to solve the problem of lack of invariance. But again, this does not mean that only models with bona fide linguistic units at the prelexical level would count as *mediated-access* models.

have to be transformed into strings of phonemes; the grouping of phonemes into syllables and the use of phonetic features have also been proposed (see Levelt, 1999, for an overview). In the speech production literature, phonemes do also often used (e.g. Guenther, 2003), but articulatory gestures have also been proposed (Browman and Goldstein, 1992). And last, but not least, syllables, phonemes and features are some of the major units of linguistic phonetic and phonological analysis – where these terms originated.

3 / A typology of word recognition models

This chapter follows directly from the last. I will consider what types of word recognition models are possible on the basis of the four descriptive dimensions introduced in Chapter 2. We will set out with a list of potential models (§3.1.1), which is then narrowed down to the types that make sense theoretically (§3.1.2 and §3.1.3). Based on the final list of model types, four questions will be identified that allow us to distinguish between them (§3.2). The question which is most central will be whether there is a prelexical level of processing.

3.1 Defining the model typology

3.1.1 A preliminary inventory of model types

In the preceding chapter we have identified four descriptive dimensions for classifying models of auditory word recognition: the three representational dimensions *abstract/concrete* (describing the complexity of a representation), *exemplar/summary* (whether the representation is open to accept new tokens, or whether it consists of a single summary description) and *structured/unstructured* (whether the representation is broken down into sub-representations); and the architectural dimension *direct/mediated* (whether there are prelexical levels of processing). Given these four dimensions, how many types of word recognition models are possible?

The abstract/concrete dimension is continuous (and it can apply at any level of processing). There are thus an infinite number of possible model types. Even if we reduced this dimension to only a small set of options, the number of possible models would still very large. For this reason I will initially omit the abstract/concrete dimension from the discussion. Once we have seen which combinations of the other three dimensions are meaningful, I will consider the abstract/concrete dimension separately for each major type.

The structured/unstructured dimension is binary. The exemplar/summary dimension can also be regarded as binary (see §2.4). The direct/mediated dimension allows us to propose as many prelexical levels as we like; but as we have seen in § 2.6 the main difference is between

models *with* and models *without* a prelexical level. We will thus treat this dimension as binary as well.

Without a prelexical level, exemplar/summary and structured/unstructured can combine in four possible ways on the lexical level:

No.	lexical representations	
1	summary	not structured
2	summary	structured
3	exemplar	not structured
4	exemplar	structured

With a prelexical level of processing, the exemplar/summary and structured/unstructured dimension can be applied to both levels. This will give use another 16 combinations (4^2):

No.	lexical representations		prelexical representations	
5	summary	not structured	summary	not structured
6	summary	not structured	summary	structured
7	summary	structured	summary	not structured
8	summary	structured	summary	structured
9	summary	not structured	exemplar	not structured
10	summary	not structured	exemplar	structured
11	summary	structured	exemplar	not structured
12	summary	structured	exemplar	structured
13	exemplar	not structured	summary	not structured
14	exemplar	not structured	summary	structured
15	exemplar	structured	summary	not structured
16	exemplar	structured	summary	structured
17	exemplar	not structured	exemplar	not structured
18	exemplar	not structured	exemplar	structured
19	exemplar	structured	exemplar	not structured
20	exemplar	structured	exemplar	structured

3.1.2 Exclusion of impossible combinations

In this section, I will reduce the total list of possible models to those that make theoretical sense. For the sake of simplicity and clarity, this processes will be illustrated with the generic model presented in §2.1. This means that (i) if a prelexical level exists, prelexical representations will be segments as in the model; and (ii) if lexical representation are structured, they will likewise be structured into segments as in the model. As mentioned repeatedly in Chapter 2, other types of representations would also be feasible.

Mixed representations will also not be considered in this model typology. This has the consequence that if I refer to a level of representation as, say, summary, I imply that there are no

exemplar representations at the same level. Finally, we will concentrate on representations required for word recognition. So if, for example, lexical representations are decomposable into some type of sublexical representation for some other reason than word recognition, we will nonetheless consider the lexicon as unstructured for our purpose.

The main principle of exclusion is straightforward: lexical and prelexical representations have to be *commensurate*. In other words, the units that make up lexical representations should of the same kind as the units that occur at the prelexical level of processing (see § 2.6).[1] For reducing our initial list of models, this principle has three applications:

1) Models with subunits at one level (in our case the lexical) that have no corresponding representations at a lower level (in our case the prelexical) can be discarded; this excludes types 2 and 4.

2) Because we only have one prelexical level in our typology, structured representations at the prelexical level are useless; this excludes types 6, 8, 10, 12, 14, 16, 18, and 20.

3) If lexical representations are unstructured, there is no reason to have a prelexical level; this excludes types 5, 6, 9, 10, 13, 14, 17, and 18.

This leaves us with the following six types:

No.	lexical representations		prelexical representations	
1	summary	not structured	-	-
3	exemplar	not structured	-	-
7	summary	structured	summary	not structured
11	summary	structured	exemplar	not structured
†15	exemplar	structured	summary	not structured
†19	exemplar	structured	exemplar	not structured

Type 15 and type 19 are problematic because they both have to be extended to models with mixed representations in order to work (see the following discussion).

3.1.3 The model typology

We will now look at all the remaining types so as to find out whether they indeed constitute meaningful possibilities. In addition, I will mention which of the models described in Chapter 1 may be subsumed under each type. For each type I will also briefly discuss the abstract/concrete dimension. It should be remembered that representations at both the prelexical and the lexical level of processing are actually pairs of representations; and abstractness of the level will be determined by the less abstract member of the pair.

[1] See the quote from Norris (1994) on page 23 about lexical representations and input representations having to be expressed in the "same vocabulary".

Type 1 – *Summary* and *unstructured* lexical representations, no prelexical representations

A model with summary and unstructured lexical but no prelexical representations is Klatt's LAFS model. Lexical representations in LAFS are summary as they consist of normalised spectral templates: each lexical entry is specified by a unique sequence of spectral templates. Even though the word templates were originally constructed from diphone templates, these subtemplates are not used on their own in word recognition. Lexical representation in LAFS are therefore not structured.

Because lexical representations have to be compared directly to the input representations in this type of model, they must be of roughly the same degree of concreteness; they have to be auditory or acoustic, in other words, depending on what we assume the input to be. LAFS uses acoustic lexical representations.

Type 3 – *Exemplar* and *unstructured* lexical representations, no prelexical representations

A model of this type is a pure multiple-trace model. Such a model is characterised by only having exemplar lexical representations and no prelexical level of processing. I call this a *pure* multiple-trace model of word recognition, because in this model words can only be recognised by comparing the input directly with the lexical traces. MINERVA 2 can be used in this way.

Regarding the abstractness of lexical representations in this model, much the same can be said as for the previous type: because lexical representations are directly compared to input representations, they need to have a comparable degree of concreteness. In addition, exemplar representations lose their raison d'être if they become too abstract: the more abstract a representation is, the less is gained by storing many of its tokens, because the tokens will all be very alike. Storage of multiple versions only make sense if the representations stored are concrete.

Type 7 – *Summary* and *structured* lexical representations, *summary* prelexical representations

Pisoni and Luce's (1987) model and the *mediated-access* models discussed in §1.2 fall into this category. In this type, word recognition is not direct but mediated by a prelexical stage of processing where subword representations (segments in our generic model) are recognised. Some of the models discussed in §1.2 have more than one prelexical level; Trace, for example, has both a feature and phoneme level.

Because there is a prelexical level of processing, lexical representations can be very abstract:

the concrete-to-abstract mapping takes place on the prelexical level. If lexical representations are structured into phonemes, for example, the phonemes must be represented at the prelexical level in a form that makes them accessible from the signal. Models with more than one prelexical level would make it possible that the concrete-to-abstract mapping takes place over several processing steps.

Type 11 – *Summary* and *structured* lexical representations, *exemplar* prelexical representations

This type is identical to type 7, except for the use of exemplar instead of summary representations at the prelexical level. A model of this type would result from the combination of a multiple-trace model of segment recognition, as e.g. proposed by Johnson (1997), with a model of auditory word recognition that takes segments as its input, such as Cohort or Shortlist.

In this type, lexical representations can again be very abstract. The only condition is that a multiple-trace recognition model can be built for the units from which lexical representations are composed.

Type 15 – *Exemplar* and *structured* lexical representations, *summary* prelexical representations

The only way this combination of lexical representations that are exemplar and structured with summary prelexical representation can possibly work is as a mixed model. The summary prelexical representations can only be used for word recognition if lexical representations were composed from commensurate sublexical units. But because lexical representations are exemplar, the lexicon would necessarily have to contain both summary and exemplar representations. Without summary lexical representations the prelexical level would be useless, and without exemplar lexical representations this type would become identical to type 7 above.

As regards the abstractness of the representations in this type, the summary sublexical representations in the lexicon can be abstract. The exemplar lexical representations, on the other hand would have to be concrete if they are to be used for word recognition.

Type 19 – *Exemplar* and *structured* lexical representations, *exemplar* prelexical representations

This last type is also problematic. Because it has exemplar lexical representations *and* a prelexical level of processing, word recognition can take place along two different paths. One path goes via the prelexical level of processing, the second path would bypass the prelexical level

and lead straight from the input to the lexicon. But then, if there are exemplar segmental representations at the prelexical level the lexicon also has to contain summary segmental units. This means that, as for type 15, this model can only make sense as a mixed model that has both summary and exemplar lexical representations.

With regard to abstractness, the same can be said as for type 15.

3.1.4 Conclusions from the model typology

We have seen that if we allow a maximum of one prelexical level of processing, there are six *possible* types of word recognition models. Two of them would only be functional with mixed lexical representations. Mixed models are not pointless, of course; on the contrary they might even turn out to be the most adequate models. But from the point of view of the experimental comparison of model types, it is best to choose models that only differ with regard to one dimension at a time. It makes sense to compare a model where lexical representations are structured with one where they are not structured. A mixed model with both structured and unstructured lexical representations should only be considered if there is strong evidence that both types of representations are required.

Disregarding the hybrid types 15 and 19 for methodological reasons, we are left with four main types of word recognition models. The model-by-model discussion of the abstract/concrete dimension has shown that, with regard to lexical representations, we can treat it as a binary dimension too, if *concrete* means 'concrete enough for a direct comparison with auditory representations', and *abstract* 'as abstract as the prelexical units allow'.

These, then, are the four categories (and the models that belong in each):

Summary direct-access model (type 1): there are *no* prelexical representations; lexical representations are *summary* and *not structured*, and they have to be *concrete* (LAFS).

Exemplar direct-access model (type 3): there are *no* prelexical representations; lexical representations are *exemplar* and *not structured*, and they have to be *concrete* (multiple trace models, such as MINERVA 2, applied to the lexicon).

Summary mediated-access model (type 7): lexical representations are *summary* and *structured*, and can be *abstract*; prelexical representations are *summary* and have to be *concrete* (Cohort, Trace, Shortlist, PARSYN).

Exemplar mediated-access model (type 11): lexical representations are *summary* and *structured*, and can be *abstract*; prelexical representations are *exemplar* and have to be *concrete* (multiple trace models applied to the prelexical level; see e.g. Goldinger, 1998).

3.2 Research questions

The four types of word recognition models can be distinguished by asking the following two questions:

1) Is there a prelexical level of processing?

2) Are lexical/prelexical representations exemplar or summary?

The first question distinguishes the two mediated- from the two direct-access types, the second the two exemplar from the two summary types. The second question can be asked of all levels of processing; and because of this the first question is arguably the more important. We first need to ask whether there is a prelexical level or not. If there is one, we should ask the second question about the prelexical level; if there is no prelexical level, we have to ask it about the lexical level.

Two questions we may also want to consider are:

3) Are lexical representations structured?

4) Are lexical representations concrete or abstract?

Question 3 may also help to distinguish mediated- from direct-access models. Mediated- but not direct-access models require that lexical representations are structured. The reverse is not true, however: direct-access models do not rule out structured lexical representations. As has I have mentioned, lexical entries may be structured into sublexical units for other reasons than word recognition. The same holds for question 4. Only mediated-access models are capable of performing word recognition with abstract lexical representations. But again, there may be other reasons why the lexicon should contain abstract representations – in addition to more concrete ones. These two questions, while important, are therefore not crucial.

There are thus three questions we could ask to distinguish *mediated-* from *direct-access* models (questions 1, 3 and 4), and one that distinguishes exemplar from summary models. Of the three questions that distinguish *mediated-* and *direct-access* models, only the first is directly about the existence of a prelexical level of processing; and it is thus the most important one. What remains to be seen is whether it is possible to distinguish this question about prelexical representations experimentally from the other two about lexical representations. In principle this should be possible, because lexical representations that are both abstract and structured may exist even without a prelexical level of processing. Questions 1 should thus be dissociable from question 3 and question 4. Whether we can find an experimental paradigm that can tell us whether there is a prelexical level of processing in auditory word recognition, I will consider in Part II.

3.3 Conclusions

The question whether a prelexical level of processing is required for auditory word recognition will be the main question of this thesis. As I hope to have made clear in this and the previous chapter a *prelexical level of processing*, as I use the term, has the specific meaning of a stage of processing where units smaller than the word are being recognised. I will only consider one kind of representation, namely the segment, because it is the prelexical representation most often used by *mediated-access* models of auditory word recognition (see §1.2).

Another thing worth remembering from the discussion in this and the previous chapter is that the two terms *abstract* and *exemplar* should not be regarded as opposites, as they refer to different representational dimensions. Abstractness is about complexity and the similarity of a representation to the acoustic signal; the term *exemplar* should be understood in opposition to *summary*, and is about different ways of storing representations. We have also seen that highly abstract exemplar representations are of little use, as storing a multitude of similar traces is only useful if they are not completely identical.

4 / Evidence for a prelexical level of processing

This chapter contains a review of existing evidence regarding the existence of a prelexical level of processing in auditory word recognition. I will present evidence from the following research areas or experimental paradigms: form priming (§4.1), phonotactic probability (§4.2), repetition priming (§4.3), perceptual learning (§4.4) and subcategorical mismatch (§4.5).

We will see that there are two studies which present good evidence in favour of a prelexical level of processing: Pallier et al. (2001), using repetition priming with Spanish-Catalan bilingual participants; and McQueen et al. (2006), using perceptual learning. The outcome of Eisner and McQueen (2005), another perceptual learning study, moreover suggests that the prelexical representations might be segmental. Some effects found in form priming studies are consistent with a prelexical level of processing, but they do not demand it. In opposition to these studies that favour *direct-access* models, Marslen-Wilson and Warren (1994) present evidence from subcategorical mismatch which suggests that, at least in some circumstances, lexical access appears to be direct.

I will draw two conclusions from this review. The first is that the currently available evidence is not yet compelling, and that further experimental studies are therefore justified. Secondly, a hybrid model that has both a *direct* and a *mediated* recognition pathway could account for all experimental findings.

4.1 Form priming

Priming will be presented in more detail and from a methodological perspective in the next chapter. For the moment, I will only look at the claim that the facilitation found in form priming, particularly with primes and targets that share stimulus-final segments, has a prelexical locus. My review of the evidence will come to a rather negative conclusion: while it is plausible to assume that some kinds of facilitative form priming has a prelexical locus, there is no compelling evidence that final-overlap facilitation has to be prelexical.

In form priming, subjects are presented with two stimuli in quick succession; the first is the prime and the second the target, and subjects only respond to the target. Primes and targets may contain identical speech sounds. What is measured is whether related primes facilitate or inhibit the processing of targets relative to unrelated primes (which share no sounds with their targets). Priming is measured both in terms of reaction time and proportion of correct responses.

In general, form priming has produced inconsistent results; they will be discussed briefly in §6.1. What is relevant to the current question is the claim that some effects occur at a prelexical level of processing. This was suggested quite early on by Slowiaczek and Hamburger (1992), who proposed that facilitation in form priming is a prelexical effect, while inhibition has its origin in the lexicon. Specifically, facilitation was hypothesised to occur when the prelexical representations used to access the target lexical representation are pre-activated by a related prime; inhibition, on the other hand, is likely to be caused by the increased competition from primed lexical representations.

The claim that facilitation and inhibition in form priming have a different cause and locus could indeed explain many of the apparent inconsistencies found in the literature. More importantly for us, it could also provide evidence for the involvement of a prelexical level of processing in auditory word recognition. Evidence in favour of a prelexical locus of certain priming effects comes from two sources: (i) priming with pseudowords, and (ii) cross-modal priming.

4.1.1 Pseudoword priming

Several studies have found that facilitation for final-overlap priming – where prime and target share several final phonemes, often their rhymes – occurs with both words and pseudowords (Slowiaczek et al., 2000, Dumay et al., 2001, Norris et al., 2002). The argument in favour of a prelexical locus of the effect goes as follows. Pseudowords do not have lexical representations; therefore, effects that occur with pseudowords as well as words should have their locus outside of the lexicon.

However, the experimental evidence from final-overlap priming is not entirely conclusive. Norris et al. (2002) found that word primes facilitate pseudoword targets in a shadowing task (where stimuli are repeated), but not in a lexical decision task (where word/non-word judgements have to be made). In Dumay et al.'s (2001) study, the facilitation from pseudoword primes was significantly larger for word targets than for pseudoword targets. These results suggest that final-overlap priming with pseudowords is not on a par with final-overlap priming with words.

In addition, only few studies have verified whether the inhibitory effects found with initial-overlap priming (where primes and targets share stimulus-initial phonemes) do not also occur with pseudowords. Radeau et al. (1989) found that, in general, initial-overlap inhibition only occurs if primes and targets are either both words or both pseudowords; but they also found some evidence that word primes may inhibit pseudoword targets. Slowiaczek and Hamburger (1992) did in general not find inhibition with initial overlap priming.[1] But they did report a non-significant increase in reaction time for word targets preceded by pseudoword primes when the amount of initial overlap was increased. These findings suggest that final-overlap priming, which is assumed to be lexical, may also occur with pseudowords.

If inhibition does indeed occur with pseudowords, we could either conclude that inhibition is non-lexical as well, or that pseudoword priming cannot, in general, be regarded as evidence for a non-lexical locus of a priming effect. As there is independent evidence that inhibitory priming has a lexical locus (Goldinger et al., 1989, Radeau et al., 1995, Dufour and Peereman, 2003), the second conclusion seems more warranted.

4.1.2 Crossmodal priming

An analogous argument can be made with regard to crossmodal priming. If a priming effect occurs when the primes are visual and the targets auditory (or vice versa), we can conclude that it cannot have a prelexical locus, because prelexical processing is assumed to be modality specific. Effects that show up in a crossmodal condition thus either have to be lexical, if we assume a modality-neutral mental lexicon,[2] or then postlexical.

Slowiaczek and Hamburger (1992) used crossmodal presentation with initial-overlap priming and found that inhibition can occur with crossmodal presentation. But in one of their crossmodal conditions they also found significant facilitation; but this was the kind of facilitation that Hamburger and Slowiaczek (1996) claim to be strategic. Dumay et al. (2001) used crossmodal presentation with final-overlap priming, and failed to find any facilitative effect. These results are consistent with the proposal that inhibition is lexical and facilitation prelexical. But these result are hardly convincing; more evidence is needed for us to firmly conclude that cross-modal priming can only produce inhibition but not facilitation, and that facilitation thus has to have a prelexical locus.

While being broadly consistent with the claim that facilitative effects have a prelexical locus,

[1] A finding which they later (Hamburger and Slowiaczek, 1996) attributed to a large amount of strategic facilitation caused by a high proportion of related trials and long inter-stimulus intervals; see §6.1.2.

[2] But consider Coleman (1998), who concludes from a review of neurological evidence that lexical representations are likely to be auditory.

the evidence from pseudoword priming and crossmodal priming does not allow us to conclude that there must be prelexical processing in auditory word recognition. In the case of crossmodal presentation, there have simply not been enough studies that have used this mode of presentation in form priming. In the case of pseudoword priming, there is more evidence available; but the evidence is inconclusive, suggesting that both facilitation and inhibition may occur with pseudowords.

4.2 Phonotactic probability

Probabilistic phonotactic facilitation – i.e. the faster processing of more common combinations of speech sounds in word recognition – has been claimed to require prelexical representations (Vitevitch and Luce, 1998, 1999, Luce and Large, 2001, Vitevitch and Luce, 2005). As with the evidence from form priming, my conclusion will be rather negative. It seems not unlikely that phonotactic properties of the speech signal may affect word recognition. The available evidence is contradictory, however, and does not warrant the conclusion that there has to be a prelexical level of processing where effects of probabilistic phonotactics occur.

Phonotactics is the study the combinatorial possibilities of the sounds of a language: which sequences of phonemes are permissible and which are not. English, for example, can only have clusters of three consonants in initial position where the first has to be /s/, the second a plosive, and the third an approximant; similarly the sequences /ps/ and /pf/ are permissible onsets in German but not in English. But phonotactics is not just a matter of what is permissible and what is not: some permissible sequences are more common than others. It has been shown that human listeners are aware of these distributional properties of sound sequences from a very early age (Jusczyk and Luce, 1994, Vitevitch et al., 1997, Gathercole et al., 1999, Frisch et al., 2000, Treiman et al., 2000, Stork, 2001, Coady and Aslin, 2004). Stimuli that have a higher phonotactic probability – i.e. that contain more common combinations of speech sounds – are preferred, and are processed faster and more accurately than stimuli with lower phonotactic probability.

Assuming that phonotactic probability may affect word recognition, we would predict that the higher the phonotactic probability of a stimulus, the faster will it be recognised. However, phonotactic probability is correlated with lexical competition: words with higher phonotactic probability (which means more common sequences of sounds) tend to be similar to more words – and thus have more potential competitors – than words with a low phonotactic probability. We also know that a larger competitor set size slows down processing (e.g. Taft and Hambly, 1986, Goldinger et al., 1989, Luce et al., 1990, Shillcock, 1990, Goldinger et al., 1992,

Norris et al., 1995, Gaskell and Marslen-Wilson, 2002, Dufour and Peereman, 2003). So here we have a potential conflict: phonotactic probability predicts facilitation where lexical competition predicts inhibition.

If we could demonstrate both effects in the same experiment, it would provide evidence that there is prelexical processing in auditory word recognition. Presumably, two opposite effects that occur simultaneously cannot originate at the same level,[3] and thus if we find evidence for their simultaneous occurrence, the two effects should have a different locus. Since we have reasons to believe that inhibition is caused by lexical competition, phonotactic facilitation has to happen outside the lexicon. Finally, if the experimental task used is an online task, the locus of the effect is likely to be prelexical – unless we have reason to believe that it might be caused not by the word recognition process but by some additional, postlexical process.

The two earliest studies that tried to demonstrate both phonotactic facilitation and lexical competition in auditory word recognition (Vitevitch and Luce, 1998 and Vitevitch and Luce, 1999) used faulty stimuli. Lipinski and Gupta (2005) discovered that the sound files used in these two studies contained some leading silence, and that the duration of the auditory stimuli was not matched across the different stimulus sets but covaried with competitor set size and phonotactic probability. This correlation of stimulus duration with phonotactic probability makes it likely that the facilitation Vitevitch and Luce found with their pseudoword stimuli was an artefact of stimulus duration, and may not have been caused by phonotactic probability.

Lipinski and Gupta (2005) investigated this problem in a series of 12 experiments, using three different stimulus sets and a variety of experimental manipulations. In the first four of their experiments they used the same pseudoword stimuli as Vitevitch and Luce (1998) and Vitevitch and Luce (1999); in the next four the same stimulus types were re-recorded without the leading silence; and in the last four experiments an entirely new set of high and low density/probability stimuli was used. Lipinski and Gupta also used two different presentation rates (either 1 or 4.5 s between stimuli), and in some experiments they digitally adjusted stimulus duration to compensate for the remaining durational differences.[4] In addition to these experimental manipulations, reaction time was measured both from stimulus onset and stimulus offset; and with the onset data, both a one-way ANOVA as well as an ANCOVA with stimulus duration as the covariate were performed.

[3] This argument is plausible; but it is only valid if we assume that each level of processing can only support one effect at a time. This is a common assumption, and models of word recognition appear to be constructed in concordance with it (in Trace, for example, connections between units on the same layer are only inhibitory), but it might not be a necessary assumption.

[4] Note that Lipinski and Gupta (2005) discovered the problem with the leading silence while carrying out the first series of experiments in an attempt to replicate the results of Vitevitch and Luce (1998) and Vitevitch and Luce (1999).

The results were as follows. In the first four experiments, with RTs measured from stimulus onset, the findings of Vitevitch and Luce were replicated: the high phonotactic probability (and large competitor set) stimuli were processed faster than the low phonotactic probability (and small competitor set) stimuli. In the ANCOVA, however, duration was the only significant factor, indicating that the original results are an artefact of stimulus duration. Measured from stimulus offset – thereby compensating for the differences in duration – the results went in the opposite direction to that reported by Vitevitch and Luce: the low phonotactic probability stimuli were responded to faster than the high phonotactic probability stimuli. In the second series of experiments, low probability stimuli were also responded to faster if measured from stimulus offset. When measured from stimulus onset, there was either no difference or responses to low probability stimuli were again faster. The third series produced a similar outcome: measured from stimulus offset, low probability stimuli were always responded to faster; measured from onset there was either no difference or still an advantage for low probability stimuli.

In none of the experiments where Lipinski and Gupta (2005) used newly prepared stimuli – which therefore did not contain the original fault – did they find a facilitative effect of competitor set size or phonotactic probability. Contrary to Vitevitch and Luce (1998) and Vitevitch and Luce (1999), this suggests that pseudowords are subject to the same inhibitory effects of lexical competition as words: pseudoword stimuli with high phonotactic probability and many lexical competitors are processed slower than matched stimuli with low phonotactic probability and few lexical competitors.

Vitevitch and Luce (2005) subsequently attempted to replicate their earlier results for pseudowords with stimulus sets that were equated for duration. They found that, measured from stimulus onset, responses to high probability/large competitor set stimuli were faster than to low probability/small competitor set stimuli, as predicted by their account of phonotactic probability. They thus managed to replicate their earlier studies with properly matched stimulus sets.

What could be the reason for the difference between the results of Lipinski and Gupta (2005) and Vitevitch and Luce (2005)? A likely candidate is a small but important difference in the experimental task. Vitevitch and Luce's subjects were given 5 seconds from stimulus onset to respond, while in Lipinski and Gupta's experiments stimuli were presented with a fixed presentation rate of 1 or 4.5 seconds. Vitevitch and Luce (2005) claim that these differences in procedure can explain the different outcomes. They tested their hypothesis for Lipinski and Gupta's shorter presentation rate, and also failed to find a significant effect of density/probability on reaction time. They suggest that the short presentation rate puts subjects under too much pres-

sure, and that the increase in error obliterates the facilitative effect of phonotactic probability.

Vitevitch and Luce (2005) did not carry out any tests with the presentation rate of 4.5 seconds, but they suggest that this longer rate may afford subjects too much time to respond, and in this way may not have produced the desired effect. Indeed, in Lipinski and Gupta's study responses are between 100 and 300 ms slower with the presentation rate of 4.5 seconds as compared to Vitevitch and Luce's study. They hypothesise that prelexical phonotactic effects may be short-lived, and has thus already dissipated by the time the responses are made.

Vitevitch and Luce's interpretation of the conflicting results is plausible, but there are nevertheless some difficulties with it. The first is that Lipinski and Gupta (2005) do not only *fail* to find facilitative effects of phonotactic probability or competitor set size, they consistently find *inhibitory* effects. These cannot be discounted as spurious. At the very least, we have to conclude that the inhibitory effect of lexical competition is considerably more robust and persistent than the facilitative effect of phonotactic probability.

Secondly, Vitevitch and Luce (2005) only report RTs from stimulus onset, while Lipinski and Gupta (2005) found consistent results for measurements taken from both stimulus onset and offset (though the effects were stronger when measured from the offset). We do not know whether Vitevitch and Luce would still have found facilitative effects if they had measured RTs from offset. Given that their high probability stimuli were 10 ms shorter than their low probability stimuli and that the facilitation they found for high probability stimuli was only 14 ms, this seems unlikely. Whether response times should be measured from stimulus offset or onset is an issue about which opinions differ; but all else being equal, an effect that does not depend on the measurement point deserves greater confidence than one that does.

From the studies discussed so far we cannot conclude that effects of phonotactic probability have been demonstrated beyond doubt. However, there was an earlier study (Luce and Large, 2001) in which phonotactic probability and competitor set size was varied orthogonally. This study was also carried out before Lipinski and Gupta (2005) discovered that Vitevitch and Luce (1998) and Vitevitch and Luce (1999) had used faulty stimuli; but because a different stimulus set was used in this study, we ought to give it the benefit of our doubt.

Competitor set size and phonotactic probability tend to covary. A lexical representations with many competitors will, all else equal, contain a more common combination of sounds than a lexical representation with only few competitors. But in spite of this correlation, Luce and Large (2001) managed to select a set of stimuli for which the two factors vary orthogonally: large competitor sets were combined with both high and low phonotactic probability, and likewise for small competitor sets. With this manipulation, Luce and Large managed to find evidence for the simultaneous presence of competition and phonotactic probability for

word stimuli. Overall, words with large competitor sets produced slower reactions than words with few competitors, and high probability words were responded faster than low-probability words. Lexical competition thus inhibits, and phonotactic probability facilitates the processing of words.

Luce and Large (2001) did not find any reliable effects for pseudowords, however. Since the phonotactic facilitation reported by Vitevitch and Luce (2005) were produced with pseudoword stimuli, this outcome deserves our attention. Luce and Large highlight that in three of their four conditions, reaction times to words and pseudowords were almost identical, while in the high phonotactic probability/small competitor set condition there is a significant difference of 30 ms. They identify the high probability/small competitor set *pseudowords* as the culprits, and suggest that for this condition their measure of lexical competition (i.e. neighbourhood density) may underestimates the true amount of competition.

This interpretation is problematic, because rather than the pseudowords in the high probability/small competitor set condition being unusual, it is the words in this condition that are anomalous. For the high probability/small competitor set words, the mean reaction time was 674 ms, while for all other conditions it was between 696 and 712 ms, and thus considerably slower. The most reasonable interpretation of this pattern is that, in general, neither phonotactic probability nor competitor set size produce an effect. Only in the most favourable condition, when there is little competition and the phonotactic probability is high, do we find significantly faster reaction times. In other words it may be the faster reaction time in the high probability/small competitor set size condition that is the root cause of both of the main effect for words reported by Luce and Large (2001).[5] If this is the case, we cannot take their results to have shown that phonotactic probability has an effect on word recognition independently of lexical competition.

In conclusion, the research into the effects of phonotactic probability on word recognition produced some evidence that phonotactic probability may facilitate the processing of words (Luce and Large, 2001) and pseudowords (Vitevitch and Luce, 2005). However, Luce and Large (2001) and Lipinski and Gupta (2005) did not find any phonotactic probability effect with pseudowords, and Luce and Large's effect for words is really only based on their high phonotactic probability/small competitor set size condition. The only firm conclusion that we can draw is

[5] You may object that we should find an interaction in this case. In theory, we should; but tests of interactions have lower power than tests of main effects. In a 2×2 design, for example, in order to demonstrate a main effect of factor A, factor B can be disregarded, so that we are in effect splitting the data into two groups or cells: A_1 and A_2. In order to demonstrate the interaction $A \times B$, on the other hand, we in effect split the data into four cells and compare whether the difference between A_1 and A_2 in B_1 is different from the difference between A_1 and A_2 in B_2. As power is dependent on sample size, testing a main effect has a higher power than testing an interaction because the test of the main effect is based on cells with larger sample sizes.

that phonotactic effects, if they exist, are shorter-lived and less robust than the effects of lexical competition.

In addition to the lack of compelling evidence, the assumption that phonotactic effects require sublexical representations may itself not be well-founded. It is true that phonotactic constraints – whether categorical or probabilistic – are most easily expressed in terms of phoneme sequences; but this does not imply that they have to be coded in this way in the speech processing system, nor that phonotactic probability needs to be computed prelexically. A lexicon containing whole words may hold all the information required: subjects' awareness of phonotactic constraints may be a form of generalisation over the whole lexicon without the need for the existence of sublexical representations.

What would be problematic, though, for a model without prelexical representations is the simultaneous occurrence of competition effects and phonotactic effects. This may be explained with just one level of processing, but an account with two levels of processing – one for each effect – seems certainly more natural. But as we have seen, it is not clear that the two effects really occur independently of each other.

4.3 Repetition priming

Two studies have used the repetition priming paradigm to address the issue of prelexical representations in auditory word recognition: Pallier et al. (2001) with a bilingual Spanish-Catalan population, and McLennan et al. (2003) using stop consonants produced in two different registers (casually and carefully spoken). The first study produces strong evidence for a prelexical level of processing; although we may question whether highly fluent bilinguals are typical. The second study presents conflicting results, but seems to favour a *direct-access* account overall.

Repetition priming, in the way I use the term, has two main features that distinguish it from form priming. First, the priming is *covert* in the sense that for the subjects there is no division of stimuli into primes and targets; from the point of view of the experimenter, however, stimuli are paired up as primes and probes. Secondly, in repetition priming there are normally *intervening stimuli* between primes and probes; primes and probes may even be presented in separate experimental blocks. More about the different priming paradigms can be found in §6.1.

4.3.1 Repetition priming in bilinguals

Pallier et al. (2001) used repetition priming – with primes and probes in the same experimental block and using a lexical decision task – in order to determine whether bilinguals have different

lexical representations depending on their dominant language. They studied two groups of highly fluent Spanish-Catalan bilinguals: Spanish-dominant and Catalan-dominant. Spanish and Catalan have many lexical items in common, but Catalan has a larger phoneme inventory. The vowels /ɛ/ and /ɔ/ and voiced fricatives, such as /z/, occur in Catalan but not Spanish. Catalan has thus minimal pairs that a Spanish does not have, such as /neta/ 'granddaughter' vs. /nɛta/ 'clean, fem.', /osos/ 'bears' vs. /ɔsos/ 'bones', and /kasa/ 'hunting' vs. /kaza/ 'house'.

Pallier et al. used three types of prime-probe pairs. Primes and probes were either *identical* (i.e. proper repetitions), *common* minimal pairs (e.g. /capa/ 'cape' vs. /cava/ 'cellar'), or *Catalan-specific* minimal pairs (e.g. /neta/ vs. /nɛta/). Both groups of bilinguals are expected to show facilitation for identical pairs, but not for the common minimal pairs because they have separate lexical representations. *Catalan-dominant* speakers are expected to treat the Catalan-specific minimal pairs exactly like the common minimal pairs. *Spanish-dominant* bilinguals may also have acquired separate lexical representations for the Catalan minimal pairs, in which case their performance would mirror that of the Catalan-dominant bilinguals. But if they have not acquired Catalan-specific representations, we expect there to occur as much priming for the Catalan-specific pairs as for the identical prime–probe pairs.

The latter is what Pallier et al. have found. Identical pairs, but not minimal pairs, produced significant facilitation in the order of 60 to 90 ms. Spanish-dominant subjects processed the Catalan-specific minimal pairs just like identical pairs, and produced a comparable amount of facilitation. These findings indicate that Spanish-dominant bilinguals indeed treat the Catalan-specific minimal pairs as homophones.

The significance of this finding for the issue of whether auditory word recognition is direct or indirect is that only models with prelexical phonemic representations predict this outcome. Catalan-specific minimal pairs, such as /osos/ vs. /ɔsos/, are treated just like identical stimuli by Spanish-dominant bilinguals because the two sounds [o] and [ɔ] both map onto /o/ prelexically, and there is only one lexical representation /osos/. A model of word recognition that does not have prelexical phonemic representations, but where the input is directly mapped to the lexicon, predicts that there should not be any priming for non-identical pairs – or at least significantly less than for phonetically identical pairs.

This repetition priming study has, to my knowledge, not been replicated yet. Of particular interest would of course be a replication with different materials and with a different population (especially a different language). But similar results have recently been reported by Sebastián-Gallés et al. (2005), again with Catalan-Spanish bilinguals (see also Sebastián-Gallés et al., 2006). They used a lexical decision task (without any priming) where the pseudowords differed from actual Catalan words by one phoneme. The change involved a contrast that was either

common to both Spanish and Catalan (e.g. the pseudoword [ʎənˈsal] derived from the word [ʎənˈsɔl] 'sheet') or specific to Catalan (e.g. the pseudoword [gəˈʎeðə] derived from [gəˈʎɛðə] 'bucket'). What Sebastián-Gallés et al. found was that Spanish-dominant bilinguals were more likely to mistake these pseudowords for words than Catalan-dominant bilinguals.

4.3.2 Priming between stylistic variants

McLennan et al. (2003) used the repetition priming paradigm with stylistic variants of words. In American English, the alveolar plosives /t/ and /d/ are both generally pronounced in intervocalic position as a voiced alveolar tap [ɾ]; *butter* is thus pronounced as [bʌɾɚ] in American English.[6] Since tapping is used with both voiceless and voiced alveolar plosives, neutralisations can occur: [ˈgreɪɾɪŋ], for instance, can either be the word *grating* /ˈgreɪtɪŋ/ or *grading* /ˈgreɪdɪŋ/. Will such an ambiguous tapped realisation prime the corresponding non-tapped realisation and vice versa? Or will tapped realisations only prime and be primed by other tapped realisations, and plosive realisations by other plosive realisations? In addition to alveolar stops, McLennan et al. (2003) used casual and careful productions of stimuli with velar (e.g. *bacon*) and bilabial (e.g. *cabin*) plosives, for which no tapping and no neutralisation occurs.

McLennan et al. (2003) presented primes and targets in two separate blocks, and used a shadowing task. Their measurement of priming was different from that used by Pallier et al. (2001). The latter subtracted RTs to probes from RTs to the corresponding primes, while McLennan et al. used measures that they call the magnitude of priming and the magnitude of specificity. Probes were primed by (i) unrelated primes in the *control* condition, (ii) phonologically and stylistically identical primes in the *match* condition, and (iii) phonologically identical but stylistically different primes in the *mismatch* condition. The *magnitude of priming* was then the mean RT to matching probes minus the mean RT to the control probes. If the reaction time to matching probes is significantly faster, we can say that priming has indeed a facilitative effect. In a case where there is priming in the match condition, the *magnitude of specificity* becomes of interest. This is the mean RT to matching probes minus the mean RT to mismatching probes. If matching probes are responded to faster than mismatching probes, we can conclude that the stylistically different variants are not treated as identical.

For stimuli containing *alveolar plosives* – for which there is neutralisation – McLennan et al. (2003) found a priming effect but no specificity effect. This means that stimuli containing tapped variants prime and are primed by stimuli containing non-tapped variants about as much as by identical stimuli. For stimuli containing *non-alveolar plosives*, both priming and

[6] This sound is often called a flap and the corresponding phonological process as flapping; however I follow Laver (1994) and Trask (1996) in using the more appropriate *tap* and *tapping*.

specificity effects were found. This suggests two things. First, alveolar plosives and non-alveolar plosive use different kinds of representations. Secondly, the results for the non-alveolar stimuli are consistent with a *direct-access* but not with a *mediated-access* account of word recognition: a *mediated-access* model with prelexical phonemic representations predicts that no specificity effect should occur – unless we want to postulate a model with as many prelexical representations as there are speaking styles. The outcome with alveolar stimuli, however, seems to be more consistent with a *mediated-access* account.

As the evidence from non-alveolar stimuli favours a *direct-access* account, we need to explain why prelexical representations seem to be used with alveolar stimuli. An obvious difference is that velar and bilabial plosives remain plosives when produced casually, whereas alveolar plosives are produced as taps. Plosives and taps are related in terms of their manner of articulation (Laver, 1994, p. 224ff., for examples, discusses taps under the heading of stop articulations); but acoustically and auditorily, the voiced alveolar tap [ɾ] is quite different from a typical voiceless alveolar plosive [t], so that alveolar plosives and taps need to be explicitly linked in some way. Prelexical representations would provide such a linkage. McLennan et al. (2003) suggest that the fact that tapping sometimes results in ambiguities may be another reason why prelexical representations are used in words containing intervocalic alveolar stops. While these prelexical representations cannot help to disambiguate the homophones, the existence of homophones may draw attention to the special status of the alveolar tap.

Another study that looked at alveolar stops is Connine (2004). Instead of using a repetition priming paradigm she used what is known as a Ganong task (after its inventor). In this task, subjects perform phonetic categorisations on acoustic continua whose end points differ in lexical status. It has been found (Ganong, 1980; see also Fox, 1984 and Pitt, 1995) that categorisation judgements are biased towards the end point of the continuum that is a word; this is also known as a *lexical bias* effect. Connine (2004) used continua such as *party–barty*, for which we expect more /p/ responses, or *better–petter*, for which we expect more /b/ responses. For each such pair she used two different continua, one where the /t/ was produced as the tap [ɾ], and another where it was produced as the plosive [t].

She found that the lexical bias effect was greater for the continuum that contained the tap. Since the locus of the lexical bias must be the lexicon, this finding shows two things. First, [t] and [ɾ] are not related to the same phoneme prelexically, or else the bias effect would have to be the same. Secondly, the lexical entry of a word such *party* must contain the tap in some form; the alveolar plosive, on the other hand, may not be encoded in the lexicon. Whatever explanation we want to give for why we get more lexical bias with stimuli containing taps, it must ultimately derive from the simple fact that the vast majority of intervocalic alveolar

plosives are produced as taps in American English (Patterson and Connine, 2001). These results are, moreover, contradicting McLennan et al. (2003), who found that alveolar taps and plosives are treated as identical.

4.3.3 Conclusion

The results from the two repetition priming studies discussed are in conflict. Pallier et al. (2001) produce convincing evidence that – at least for their Spanish-dominant Spanish-Catalan bilinguals – prelexical phonemic representations are likely to be used in auditory word recognition. What remains an open question, is whether this finding generalises to other populations, even to other bilingual populations. McLennan et al.'s (2003) findings suggest that – apart for the special case of intervocalic alveolar stops – phonemes produced in different speaking styles are not equivalent; this is more consistent with a *direct-access* model. In the case of alveolar plosives, however, tapped and non-tapped productions are treated as equivalent; this is more consistent with a *mediated-access* account. This result seems slightly doubtful, if we consider Connine's (2004) finding that the Ganong effect is stronger for tapped than non-tapped productions.

Unless we want to claim that Pallier et al.'s results for bilinguals are somehow anomalous, the data from repetition priming favours a hybrid model, that has both a direct and an indirect route from the input to the lexicon. In addition, the representations used at the prelexical level appear to be phonemes, or at least segments.

4.4 Perceptual learning

Norris et al. (2003), Eisner and McQueen (2005), McQueen and Mitterer (2005) and McQueen et al. (2006) used the lexical bias effect to alter some of the lexical representations of their subjects. They then verified, among other things, whether this learning occurred at a prelexical level. McQueen et al. (2006) present the clearest evidence to date for the necessity of prelexical processing in auditory word recognition. Eisner and McQueen (2005) present evidence to suggest that the scope of the effect may be segmental.

The paradigm, first presented by Norris et al. (2003), works as follows. Subjects perform a lexical decision task on a set of words and pseudowords. Some of the words contain an ambiguous fricative that, in a pretest, has been shown to be perceived as halfway between a typical Dutch /s/ and /f/. (All experiments have been carried out in Dutch.) One group of subjects – we may call them the [s]-*ambiguous* group – hears the ambiguous fricative in stimuli that are Dutch words only if the ambiguous sound is interpreted as an /s/ (as in e.g. *naadelbos* 'pine

forest'); all words ending in /f/ contain unambiguous labiodental fricatives. The [f]-*ambiguous* group hears the ambiguous sound in stimuli that are words only when the ambiguous sound is treated as an /f/ (as in e.g. *witlof* 'chicory'); all occurrences of /s/ are unambiguous.

The point of this manipulation is to make the [s]-*ambiguous* group treat the ambiguous sound as a, somewhat unusual, variant of a Dutch /s/, while the [f]-*ambiguous* group will treat it as a, again unusual, variant of /f/. Whether this is the case can be tested with either a phonetic categorisation task (Norris et al., 2003, Eisner and McQueen, 2005, McQueen and Mitterer, 2005) or a priming task (McQueen et al., 2006). By using test stimuli which subjects have not heard in the training, one can also find out whether the learning generalises to new stimuli and what the condition of the generalisation are.

Norris et al. (2003) used a four-steps [ɛf–ɛs] continuum in their categorisation task. The [f]-*ambiguous* group categorised significantly more tokens along the continuum as containing an /f/ than did the [s]-*ambiguous* group. These results were compared with the performance of several control group, to make sure that the observed shift in the category boundary was indeed caused by a lexical bias and not a merely perceptual effect, such as selective adaptation.

Eisner and McQueen (2005) extended these findings by varying the identity of the speaker between training and test. In their baseline experiment, they used the same (female) speaker in the lexical decision and the phonetic categorisation task, and replicated the shift in the category boundary found by Norris et al. (2003). They then changed the vowel in the [ɛf–ɛs] continuum, so that it was produced by a different speaker; the fricatives were the original ones. The category shift was still observed, regardless of whether the second speaker was female (as was the original speaker) or male. If, however, the whole [ɛf–ɛs] continuum was produced by the male speaker, no shift in the category boundary was observed. If the fricatives in the training stimuli – *only* the fricatives – were produced by the same male speaker who produced the continuum, a reliable boundary shift was again observed.

We can draw two conclusions from these results: (i) that perceptual learning is speaker-specific; and (ii) that its scope is the segment. The lexical bias induced in the lexical decision task only results in a boundary shift in the subsequent phonetic categorisation if the fricative segment is produced by the same speaker in both tasks. If the phonetic context is changed from training to test the effect still occurs, but not if the fricative itself is changed.

McQueen et al. (2006) took this argument in favour of prelexical segmental representations one step further. Norris et al. (2003) and Eisner and McQueen (2005) only showed that adjustments to unusual pronunciations have a segmental scope, but not that they occur at a prelexical level of processing. The phonetic categorisation paradigm used in these studies does not allow us to conclude that the locus of the adjustments is prelexical, because it is a metalinguistic task

that may be performed either postlexically (i.e. after the words have been recognised) or outside the word recognition process. McQueen et al., therefore, used a different task, one that is much more likely to be informative about the lexical activation process itself: crossmodal priming with a lexical decision task and an intra-stimulus interval of zero milliseconds.

The training was identical to Norris et al. (2003). In the test, subjects had to perform lexical decisions on visual targets preceded by auditory primes. The targets were Dutch minimal pairs, such as *doof* 'deaf' and *doos* 'box'. These targets were preceded either by *related* but ambiguous spoken primes (*doof/doos* but with the ambiguous fricative used in the training) or by *unrelated* primes (*krop* in the present case). Note that the words used in the test were novel words which subjects had never before heard pronounced with the ambiguous fricative. Would these related but ambiguous primes facilitate the processing of the targets relative to the unrelated primes? If there was facilitation, then McQueen et al. (2006) predicted that it would only occur if the visual target ended in the consonant which the subject had heard in the ambiguous form during the training task: the [f]-*ambiguous* group should show facilitation with related ambiguous primes if the target ended in /f/, but not if it ended in /s/, and vice versa for the [s]-*ambiguous* group.

For the [f]-*ambiguous* group, the results were entirely as predicted: responses to visual targets ending in 'f' were facilitated when preceded by a related prime, and for targets ending in 's' there was a non-significant inhibition. The outcome for the [s]-*ambiguous* group was inconclusive. McQueen et al. suggest this is because all the ambiguous stimuli used in the training task were derived from recordings of /f/-final words; they are thus likely to contain acoustic cues which are inconsistent with the interpretation of the ambiguous fricative as being a realisation of /s/.[7]

McQueen et al. (2006)'s findings for the [f]-*ambiguous* training group indicate that prelexical segmental representations are involved in auditory word recognition. The ambiguous fricatives were treated by this group as realisations of /f/, or else there would have been no facilitation of visual stimuli containing the letter 'f'. And because subjects had never before heard the test words with the ambiguous fricatives, the most plausible explanation is that the perceptual learning did not affect lexical representations, but the prelexical representation of /f/. A direct *access-model* seems incapable of explaining this outcome, because adjustments to individual lexical representations cannot explain the generalisation of the perceptual learning to novel stimuli.[8]

Taken together the studies using the *perceptual learning* paradigm provide the strongest evidence yet that auditory word recognition requires segmental sublexical representations. The

[7] Note that Norris et al. (2003), who used the same training stimuli, had already noted a tendency to interpret the ambiguous fricatives as /f/.

[8] McQueen et al. (2006) refer to a study (not yet published at the time of writing) showing that a MINERVA 2-type model cannot predict these results.

experiments of Eisner and McQueen (2005) shows that when listeners have to deal with an atypical pronunciation of a speech sound they make narrowly localised adjustments that are moreover speaker-specific. The fact that they are speaker-specific suggests that representations are quite detailed and include information about the identity of the speaker, but the narrow localisation indicates that these detailed representations are not representations of whole words, but are composed of smaller units which may have the size of the traditional segment. But because a phonetic categorisation task was used in these experiments, we cannot conclude that these segmental units function at a prelexical stage in word recognition – they could be postlexical.

McQueen et al. (2006) improved on these results in two ways. First by using a task that implies a prelexical locus, and secondly by showing that the adjustment made during training generalise to words not encountered in the training, an outcome which is difficult – if not impossible – to explain for a *direct access* model.

One could argue that a sound halfway between /f/ and /s/ is quite unnatural, since the two fricatives have a place of articulation that involve different articulators: lower lip against upper teeth for the labiodental /f/, and tongue blade against alveolar ridge for the alveolar /s/. A pairing of /s/ with either /θ/ or /ʃ/ would seem more natural. I do not think that this issue of naturalness is decisive – after all, most subjects had no problems treating the ambiguous fricative as either an /f/ or an /s/, though there was a much higher rate of non-word responses to the ambiguous s-words than to any other type of training words – in all experiments, even if the ambiguous stimuli where not based on /f/-final words. In any case, it would be a desirable to replicate the experiment of McQueen et al. (2006) with a different ambiguous sound and with language other than Dutch. McQueen and Mitterer (2005) used the perceptual learning paradigm with vowels, but they used phonetic categorisation as their test task. What is needed is a replication using the crossmodal priming paradigm or another online task.

4.5 Subcategorical mismatch

When a stimulus contains conflicting cues to the identity of one of its segments – e.g. one cue suggesting that it is a /p/, the other that it is a /t/ – where will this conflict be resolved? In the lexicon as a *direct-access* model would predict, or at a prelexical level as predicted by a *mediated-access* model? Several studies have used *subcategorical* mismatch to investigate this issue. The results are not clear-cut; however, they seem to favour a *direct-access* account or a *hybrid* model with both a direct and mediated access route.

In the following I will first survey the experimental data and then consider a computer sim-

ulation of that data.

4.5.1 Empirical evidence

Marslen-Wilson and Warren (1994) cross-spliced stimuli so that they contained conflicting cues to the place of articulation of the final consonant. For example, the word *job* was cross-spliced from the onset and nucleus of /dʒɒg/ and the coda of /dʒɒb/. It is well known that the acoustic cues to obstruent place of articulation are located partly in the obstruents themselves – for plosives mainly the release burst, and for fricatives the frequency distribution of the friction noise – but also in the preceding vowel, particularly the formant transitions (see Hayward, 2000, pp. 174–207, or any other phonetics textbook). The cross-splicing of /dʒɒg/ with /dʒɒb/ thus produces stimuli with conflicting cues to the place of articulation of the final obstruent: the transitional cues in the vowel suggest a /g/ and the word *jog*, but the release burst will make it apparent that the final consonant is a /b/, and that the word is *job*. Such conflicting cues – also called a *subcategorical mismatch* – have been shown to inhibit processing relative to stimuli without conflicting cues (Streeter and Nigro, 1979, Whalen, 1984).

The stimuli that Marslen-Wilson and Warren (1994) used were not all based on words, as in the example above; they could also be constructed from a word and a pseudoword or from two pseudowords. Whether a mixed-lexicality stimulus will be heard as a word or a pseudoword depends on the second cross-spliced component, as it is the consonantal cues that tend to determine the final percept. Three different versions of each type of word or pseudoword were produced. A word such as /dʒɒb/ was created from (i) two realisations of /dʒɒb/ itself, (ii) from the words /dʒɒg/ + /dʒɒb/, or (iii) from the pseudoword /dʒɒd/ and the word /dʒɒb/. A pseudoword such as /smɒb/ was created from (i) two realisations of /smɒb/ itself, (ii) from the word /smɒg/ and the pseudoword /smɒb/, or (iii) from the pseudowords /smɒd/ + /smɒb/. The non-mismatching stimuli form the baseline relative to which inhibition is measured. Note that of the other two stimuli, one is made from components with equal lexicality (both words or both pseudowords) and the other from components that differ in lexicality. See Table 4.1 for a more detailed overview of the stimuli. Phonetically, the stimuli in each such triplet differed in place of articulation only, and the final obstruents used were the voiced plosives /b, d, g/, the voiceless plosives /p, t, k/, the voiced fricatives /v, z, ʒ, ð/, and the voiceless fricatives /f, s, ʃ, θ/.

This use of stimuli with components of different lexicality allowed Marslen-Wilson and Warren to test the different predictions of *mediated-* and *direct-access* models. A *mediated-access* model with segmental prelexical representations predicts that inhibition should occur regardless of the lexical status of the components, because the integration of the conflicting information is predicted to take place at the prelexical stage of processing. On a *direct-access* account –

Type	Composition	Example	Percept	Mismatch	MA pred.	DA pred.
W1W1	word 1 + word 1	/dʒɒb + dʒɒb/	/dʒɒb/	no	baseline	baseline
W2W1	word 2 + word 1	/dʒɒg + dʒɒb/	/dʒɒb/	yes	inhibition	inhibition
N3W1	nonword 3[9]+ word 1	/dʒɒd + dʒɒb/	/dʒɒb/	yes	inhibition	inhibition
N1N1	nonword 1 + nonword 1	/smɒb + smob/	/smɒb/	no	baseline	baseline
W2N1	word 2 + nonword 1	/smɒg + smob/	/smɒb/	yes	inhibition	inhibition
N3N1	nonword 3 + nonword 1	/smɒd + smob/	/smɒb/	yes	inhibition	no inhibition

TABLE 4.1: Example stimuli from the subcategorical mismatch experiments reported by Marslen-Wilson and Warren (1994) and McQueen et al. (1999), with predictions for *mediated-* and *direct-access* models. Note that the only difference is in the nonword+nonword condition (N3N1) in the last row. The results of the experiments are shown in FIGURE 4.1. (Adapted from Marslen-Wilson and Warren, 1994.)

because the integration of the conflicting information has to take place in the lexicon itself – a mismatch should only result in inhibition if at least one of the components is a word. For the third type of pseudoword stimulus (N3N1: /smɒb/ made from the two pseudowords /smɔd/ + /smɔb/) a *direct-access* model predicts no inhibition while a *mediated-access* model predicts inhibition (see TABLE 4.1). The predictions for the word stimuli are the same for both models; they can thus be taken as a test of the *subcategorical mismatch* paradigm.

Marslen-Wilson and Warren (1994) found reliable inhibition due to subcategorical mismatch only for stimuli ending in voiced plosives. Stimuli ending in fricatives did not produce any inhibition at all. Stimuli with voiceless plosives produced a small, but non-significant inhibition with the word stimuli and none with the pseudoword stimuli.[10] These stimuli could thus not be used to test the main hypothesis, as the occurrence of inhibition is its prerequisite; thus the following discussion is only about the stimuli ending in voiced plosives.

Marslen-Wilson and Warren (1994) used lexical decision and phoneme identification as their test tasks. In the lexical decision experiment, mismatching words reliably produced inhibition. Mismatching pseudowords only resulted in a significant inhibition if the vocalic cues suggested a word (W2N1) but not if they suggested a pseudoword (N3N1), as predicted by the *direct-access* account. In the phoneme identification experiment, Marslen-Wilson and Warren found a significant inhibition also for the stimuli created from two pseudowords (N3N1), but it was also significantly smaller than the inhibition produced by mixed-lexicality pseudowords (W2N1) and the mismatching words. The reaction time results of both experiments are shown

[9] Why there is a word 2 but a nonword 3, I do not know; this is the notation used by Marslen-Wilson and Warren (1994) and all subsequent studies.

[10] The explanation Marslen-Wilson and Warren (1994, p. 657) give for why mismatching transitions do not inhibit the processing of fricatives and voiceless plosives is that for these sounds the place information of the consonantal cues (friction and stop release) is so dominant that the mismatching transitional information is of no consequence.

in FIGURE 4.1 (broken lines).

These results indicate that – at least in the case of final voiced plosives – the conflicting information from the transition and the release burst gets resolved in the lexicon. If the conflict were resolved prelexically, the inhibition caused by mismatching cues should not depend on the lexical status of the components that provide these cues. Marslen-Wilson and Warren's evidence is thus consistent with *direct-access* and inconsistent with *mediated-access* models. It does not, however, show that there cannot be any prelexical phonemic representations in auditory word recognition, but simply that there has to be a way in which sub-phonemic information can bypasses these phonemic representations and directly influence lexical processing. A hybrid model could thus also account for their data.

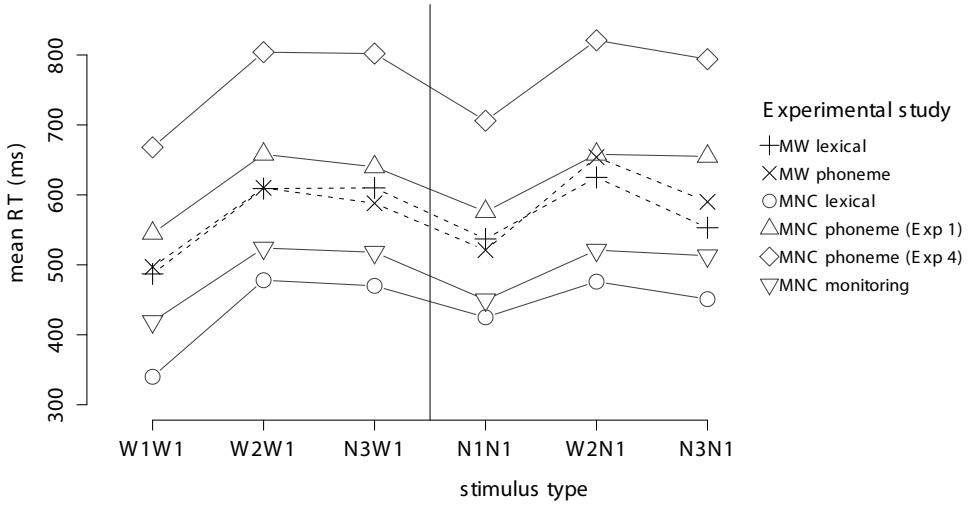


FIGURE 4.1: Overview of the subcategorical mismatch experiments: RTs for the stimulus types in the different experiments. MW (dashed lines) stands for Marslen-Wilson and Warren (1994), MNC (solid lines) for McQueen et al. (1999); *lexical* refers to the lexical decision task, *phoneme* to the phoneme identification task, and *monitoring* to the phoneme monitoring task. The stimulus types on the x-axis are described in TABLE 4.1.

McQueen et al. (1999) replicated and extended the findings of Marslen-Wilson and Warren (1994) with Dutch stimuli. There main focus was not auditory word recognition but different theories of phonetic decision making, but their results nonetheless speak to the same issues as Marslen-Wilson and Warren's study. As Dutch does not have any word-final voiced plosives, McQueen et al. carried out their replication with stimuli ending in voiceless plosives. They managed to replicate Marslen-Wilson and Warren's findings with the lexical decision paradigm

('MNC lexical' in FIGURE 4.1), despite using voiceless plosives.[11] But with the phoneme identification paradigm ('MNC phoneme (Exp 1)') and with a phoneme monitoring task ('MNC monitoring'), they found that inhibition may also occur for stimuli where both the transitional and the stop release information suggest pseudowords, favouring a prelexical locus of integration. But by adding filler items that ended in fricatives and nasals ('MNC phoneme (Exp 4)'), they managed to replicated Marslen-Wilson and Warren's results with the phoneme identification task, though there was still a large amount of inhibition with the N3N1 stimuli (see FIGURE 4.1).

Dahan et al. (2001) used subcategorical mismatch stimuli in an eye-tracking experiment with the visual world paradigm introduced by Allopenna et al. (1998). Their study focused on the issue of lateral inhibition in lexical competition, and for this reason they only used the word stimuli W1W1, W2W1 and N3W1. Because they did not consider the crucial N3N1 case, their experimental results do, unfortunately, not speak to the issue of whether word recognition is direct or mediated. It is interesting to note, however, that they conclude that eye-tracking provides a more fine-grained measure of lexical competition – that in addition sheds light on the time course of processing – than the lexical decision paradigm used by Marslen-Wilson and Warren (1994) and McQueen et al. (1999). Eye-tracking applied to subcategorical mismatch should thus be an avenue worth pursuing.

The combined data of the Marslen-Wilson and Warren and McQueen et al. studies appears not to be consistent with either a *direct-* or a *mediated-access* model. They could be explained by a *hybrid* model or by a *direct-access* model where part of the phoneme identification and monitoring results from a *postlexical* process.

The general pattern of all the experiments is very similar apart from the large overall differences in reaction time, as can be seen in FIGURE 4.1. With regard to the crucial N3N1 condition, there is one experiment ('MW lexical') in which there was no inhibition whatsoever, strongly supporting the *direct-access* account where the mismatch is resolved in the lexicon. In three of the experiments ('MW phoneme', 'MNC lexical', and 'MNC phoneme (Exp 4)') there was a significant inhibition in the N3N1 condition, but the amount of inhibition was also significantly smaller than in the conditions where at least one of the cross-spliced items was a word; these results are less clear-cut, but they can nonetheless be taken to support the *direct-access* account, since at least some of the mismatching information has to be combined in the lexicon in order to account for the difference between N3N1 and W2N1. Two of the six experiments support a *mediated-access* account: in both 'MNC phoneme (Exp 1)' and 'MNC monitoring', there was as much inhibition in the N3N1 as in the W2N1 condition, suggesting that the mismatch is resolved

[11] This may indicate that the relative weight given to the vocalic and consonantal cues to stop consonant place of articulation is language-specific, with more weight being assigned to the transitional information in Dutch than in English. There may, of course, be other reasons for this difference.

prelexically.

These conflicting experimental data could be explained a *hybrid* model. This model would have two integration sites for the conflicting acoustic information: a lexical site, which appears to be the main integration site for the lexical decision task (and may also be used in other tasks that foreground lexical processing); and a prelexical site, which appears to be used in task such as phoneme identification and monitoring which draw attention to sublexical units. If such a hybrid model is the most adequate, we have to address the issue of how it copes with everyday word recognition, i.e. word recognition that is not affected by the additional demands of experimental tasks: will it be mainly direct or mediated, or will both 'access routes' be open at all times?

Another model that could presumably account for the existing data is a *direct-access* model where phoneme identification and monitoring would involve a postlexical stage of processing. This would explain why the difference between the W2N1 condition (inhibition) and the N3N1 (no or little inhibition) seems to occur mainly in the lexical decision tasks. It might also explain why RTs in the lexical decision task were faster than in the other two tasks, particularly in McQueen et al.'s 1999 study; though there may, of course, be other explanations for this fact.

4.5.2 Computer simulations

Before closing this section on the subcategorical mismatch paradigm, let us look at a simulation of the experimental data reported by Norris et al. (2000). They show that Merge – a model of phonemic decision making based on Shortlist (Norris, 1994), see FIGURE 4.2 – can simulate the experimental data of Marslen-Wilson and Warren (1994) and McQueen et al. (1999). Because Merge has a layer of phoneme nodes, Norris et al. claim that these simulations show that the experimental data does not speak against models with prelexical representations, contrary to Marslen-Wilson and Warren's conclusion. This claim is somewhat problematic, for two reasons. First, it seems debatable whether Norris et al.'s simulations really capture the experimental data. Secondly, what kind of model, in my model typology, is Merge? I wish to argue that, even though Merge has a prelexical stage of processing, it should not be regarded as a *mediated-access* model.

First, can we truly say that Norris et al. (2000) managed to model the experimental data, as they claim to have done? Several commentators have pointed out that their choice of an activation threshold of 0.20 in the simulation of the lexical decision data is arbitrary and lacks independent support (Tanenhaus et al., 2000, p. 349); and that there is only a very narrow window for the activation threshold (between 0.18 and 0.24) within which the experimental data is adequately simulated (Gaskell, 2000, p. 330). Independent support for the parameter

settings is obviously desirable, but we cannot reject simulations out of hand simply because their parameter settings lack in support. In this particular case, however, where the value of the threshold is so crucial to the outcome of the simulation, we require an explanation for why it has to be 0.20 and not, say, 0.25. And in the absence of an explanation, we should at least ask for a demonstration that the same value of 0.20 can also be used to simulate other speech perception behaviour; should different data require wildly different activation thresholds, the argumentative force of the simulation is much reduced.

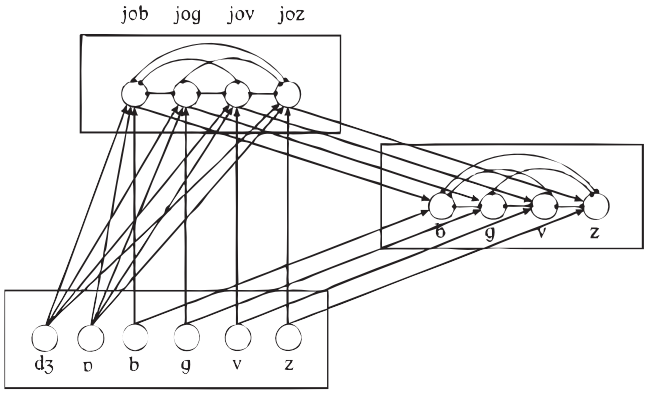


FIGURE 4.2: The *Merge* model. The model contains three types of nodes: phoneme input nodes are shown on the bottom left; they are connected to the lexical node (top) and the phoneme decision nodes (right). The phoneme decision level merges information – hence the name of the model – flowing from both the phoneme input nodes and the lexical nodes. Note that there are no inhibitory connections between phoneme nodes; there is thus no competition and also no decision process on this stage of processing (see running text for further discussion). Taken from Norris et al. (2000, p. 313).

Another problem with Norris et al.'s simulation is, I think, the nature of the input. The input was always /dʒɒb/, simulated on the phonemic input layer by the /dʒ/ node becoming activated at time slice 1, the /ɒ/ node at slice 4, and the /b/ node at slice 7. In the non-mismatching case the activation in these nodes would build up over three time slices, from an initial level of 0.25 to 0.5 to a final level of 1.0; the cross-splicing was simulated by giving the /g/ node an activation value of 0.15 from slice 7 onward, and giving the /b/ node successive values of 0.25, 0.5 and 0.85 (instead of 1.00 in the non-mismatching case). The different types of cross-spliced stimuli used in the experiments were simulated not by different inputs, but by the lexical nodes that were included in a simulation. To simulate the W1W1 and W2W1 conditions, for example, only *job* and *jog* were included; for the crucial N3N1 condition the nodes were *jov* and *joz*. In other words, the lexical status of the stimuli was not simulated on the input level but via the use of either word or pseudoword nodes on the lexical level.

While I am not claiming to be an expert on neural network modelling, there seems to me

some doubt about whether Norris et al. (2000) can claim to actually model the experimental data, where the difference occurred in the acoustic input. Another, less crucial, issue regards the simulation of the mismatch. In the simulations, the conflicting information was made available at the same time, with the /g/ node always being activated less than the /b/ node, while in the experiment the acoustic information favouring /g/ arguably becomes available earlier (in the transition as opposed to the release burst) and is initially stronger than the information favouring /b/.

Given all these issues with Norris et al.'s 2000 simulations, it is not at all clear whether they really have demonstrated that Merge can model the experimental data. And even if they have, my second point is that the Merge model as used in their simulations should either be classified as a form of *direct-access* model with a *postlexical* phoneme recognition module, or else as a *mediated-access* model whose prelexical level is *featural* and not phonemic (again with a *postlexical* phoneme recognition module).

My main reason for making this claim is that in Merge no phonemic decisions are made at the input phoneme level; these decisions are made at a later stage that merges the information from both the phoneme stage and lexical stage. The input phoneme stage is thus not a *level* in my sense of the word, since no decisions are made at this stage of processing: the phoneme nodes simply pass on the conflicting information they receive from the cross-spliced stimuli to the lexical level. In short, whether a /b/ or /g/ is present in the input is not decided at the prelexical stage, but either at the lexical level (in lexical decision) or at the postlexical phoneme decision level (in phonetic categorisation, phoneme identification or phoneme monitoring). Depending on what we assume the input to the model to be, this characteristics is consistent with either a *direct-access* model where the input is completely unclassified or with a *mediate-access* model that has a feature level prior to the phoneme stage of Merge.

To sum up this discussion of the subcategorical mismatch data, the evidence is equivocal but seems to favour a *direct-access* account on balance. However, to explain the data from the phoneme identification and monitoring task, the direct access route needs to be augmented by a mediated one, which would result in a *hybrid* model. Alternatively we might want to conclude that the phoneme identification and monitoring data are the result or a postlexical processing module and do, therefore, not directly reflect auditory word recognition.

4.6 Conclusions

The studies reviewed in this chapter present evidence both for and against the involvement of a prelexical stage of processing in auditory word recognition.

Pallier et al.'s (2001) repetition priming experiment on Spanish-Catalan bilinguals and the perceptual learning paradigm used by Norris et al. (2003), Eisner and McQueen (2005) and especially McQueen et al. (2006) produced the most persuasive evidence that auditory word recognition involves a prelexical level of processing. Eisner and McQueen's findings in addition suggest that the sublexical representations used at this prelexical level could be segments.

This prelexical stage of processing could also be the place where some form priming effects and effects of probabilistic phonotactics occur. As we have seen in §4.1 and §4.2, however, the evidence is not compelling in these two cases. Facilitative effects of form priming are well established, particularly with final-overlap priming, but the evidence that they have to have prelexical locus is much weaker. With regard to probabilistic phonotactics we have to conclude that, at present, the data is contradictory, and it is unclear whether probabilistic phonotactics affects auditory word recognition at all.

Other studies – McLennan et al., 2003 and Connine, 2004 looking at allophonic variation, and Marslen-Wilson and Warren, 1994 using stimuli with mismatching phonetic cues – have found that in some cases access to the lexicon may be direct and without recourse to prelexical representations. Studies of speaker variation indicate that lexical representations must contain quite detailed information about speaker identity (e.g. Craik and Kirsner, 1974, Schacter, 1992, Palmeri et al., 1993, Church and Schacter, 1994, Goldinger, 1996). These results are also consistent with the hypothesis that lexical access is direct; but as mentioned earlier (§1.3.2), they do not prove that lexical access has to be direct.

The existing research literature thus provides evidence consistent with both *direct-* and *mediated-access* models. Either some of the reported findings are spurious or have another explanation, or we have to conclude that an adequate model of auditory word recognition should be a *hybrid* model. It should have a prelexical level of processing where segments are recognised; and consequently, lexical entries should also consist of segments. But at the same time the lexical representations need to contain some more fine-grained acoustic information than mere strings of phonemes can contain; and there must be an alternative passageway through which information may percolate up to the lexicon.

There are plenty of unresolved issues still. The first is that the evidence for either type of model is not overwhelming. The study by Pallier et al. (2001) has not been replicated yet. In the case of Marslen-Wilson and Warren's (1994) evidence for direct lexical access, we have seen that their results have been qualified and the thrust of their argument weakened by McQueen et al. (1999) and the simulations of Norris et al. (2000). The perceptual learning paradigm has withstood several replications (Norris et al., 2003, Eisner and McQueen, 2005, McQueen and Mitterer, 2005, McQueen et al., 2006), but only the most recent of these studies has reported

evidence that strongly speaks in favour of prelexical processing. Additional evidence seems required still.

Part II

The experimental study

In Chapter 3, I have argued that the the fourth of the descriptive dimensions – whether a pre-lexical level of processing is required for auditory word recognition – is crucial if we want to distinguish between types of word recognition models. This is the question that I have addressed experimentally; and the outcome of this experiment will be presented in this second part of the thesis.

The experiment has a relatively complex structure. It consists of two training sessions, and one test session with two different test tasks: repetition priming and phonetic categorisation. To make my presentation easier to follow, I will first present the basic design and discus what is needed to implement the design (Chapter 5). Then I will briefly review some of the literature about the tasks chosen (Chapter 6). This chapter is primarily intended as an introduction for those not familiar with the tasks; but it will also serve as a justification of why I have chosen repetition priming and phonetic categorisation as test tasks. Then, in Chapter 7, I will describe how I have decided to implement the design. Chapter 8 states the predictions that *direct-* and *mediated-access* models make regarding the two test tasks. The last three chapters present the results of the experiment: Chapter 9 for the training, Chapter 10 for the repetition priming task, and Chapter 11 for the phonetic categorisation task.

5 / Design

The main research question – whether a prelexical level of processing is necessary for auditory word recognition – can be formulated in terms of two competing types of models, which in Chapter 2 I have introduced as *direct-* and *mediated-access* models. The two types are illustrated in FIGURE 5.1 (repeated from §2.6, p. 47).

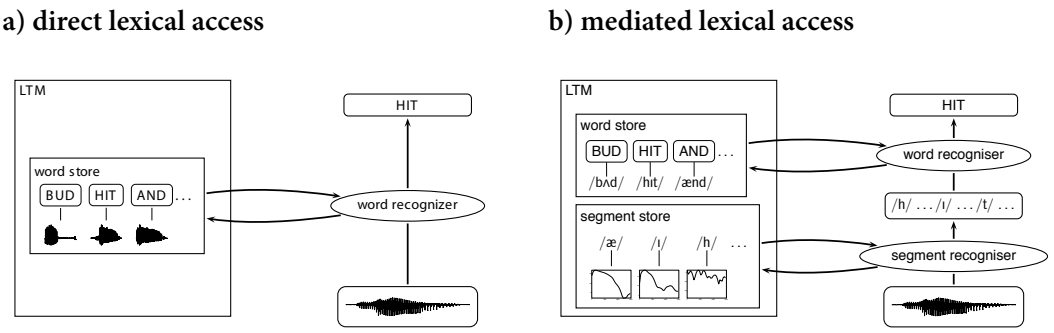


FIGURE 5.1: Direct- and mediated-access models. LTM stands for long-term memory. Square boxes represent memory stores, ellipses processing mechanisms, and boxes with rounded corners represent the objects on which the processing is carried out.

Notice the main differences between the two types. According to *direct-access* models, word recognition is a one-stage process where access is made directly from a suitably pre-processed speech signal to the lexicon. *Mediated-access* models propose that word recognition is a two-stage process where the recognition of words involves the recognition of smaller units. These smaller units could be of various size; but I will focus on the segment because most models that belong in the *mediated-access* category at the least recognise some form of segmental prelexical representations (see §1.2).[1] This difference in processing between the two types is mirrored by a difference in what needs to be stored in long-term memory. *Direct-access* models only

[1] The term *segment*, as used here, refers to the smallest *temporal* unit of speech. Crucially, it is not specified for its level of abstraction; the term *segment* thus encompasses both phonemes and phones. *Phonemes* are phonological segments, and *phones* are phonetic segments. The term *allophone* is also used to refer to phonetic segments, but as mentioned earlier it is generally reserved for the more specific case where two or more phones are realisations of the same phoneme.

require whole words to be stored, *mediated-access* models require the storage of both lexical and sublexical representations – words and segments in our case.

The difference in storage has obvious consequences for learning. In both types, new items can be added to the lexical store, but only in *mediated-access* models can learning occur in the segment store as well. This difference in learning is pivotal for my experiment. I will train subjects to recognise a set of new words that contain a non-native (i.e. non-English) speech sound; this sound is the voiceless bilabial fricative [ɸ]. On a *mediated-access* account, these new words can only be recognised successfully if subjects have acquired a segmental representation for the new speech sound *in addition* to the lexical representations for the new words. In contrast to this, *direct-access* models predict that recognition can be successful without the formation of segmental representations.

To test these different predictions regarding learning we need to do the following. After subjects have shown themselves able to recognise the new words, we have to find out whether by acquiring them they have also acquired segmental representations for the non-native speech sound [ɸ]. Two components are thus needed to carry out the experimental study as proposed:

1) an effective way of training subjects to recognise the set of new words; and

2) a procedure that allows us to determine whether segmental representations exist for a speech sound.

The first component is mainly a matter of practicality: can we find a training procedure that is suitably efficient and reliable? The second is less trivial. As we have seen in our review of previous research in Chapter 4, it is debatable what may count as legitimate evidence for the existence of a perceptual unit in a given theoretical context and experimental condition. I have identified two procedures which – combined with the training procedure – should be capable of providing sufficient evidence to answer our research question: repetition priming and phonetic categorisation.

The most important ingredient of the experimental study is, however, the training procedure. The experiment produces outcomes that I think are germane to auditory word recognition *because* subjects acquire the non-native sound [ɸ] by being trained to recognise words that contain it.

In the remainder of this chapter, I will briefly describe the training and test tasks. Some background to the experimental methods will be presented in Chapter 6, and Chapter 7 will provide a more detailed description of the experimental tasks and procedures.

5.1 The training task

In two training sessions, participants were exposed to a set of four stimuli, in the form of two minimal pairs. The meaning of each stimulus was given by an image. FIGURE 5.2 presents all auditory and visual training stimuli. Subjects had to learn to associate each word with its corresponding image; for this purpose, they received training in recognising and distinguishing the words.

a) phonemic training group

b) allophonic training group

FIGURE 5.2: Minimal pairs used in the training sessions, with the image specifying their meaning. The PHONEMIC training group will hear the [f]- and [ɸ]-stimuli with different images; the ALLOPHONIC group will hear them with the same image.

There were two different training groups. The PHONEMIC group had to learn the four words /tɪnˈdef/, /tɪnˈdeɸ/, /pəˈkif/ and /pəˈkiɸ/, where each word came with a different image. Participants were taught in this way to regard /tɪnˈdef/ vs. /tɪnˈdeɸ/, and /pəˈkif/ vs. /pəˈkiɸ/ as minimal pairs, and to treat the non-English bilabial–labiodental fricative contrast as a *distinctive* contrast. The ALLOPHONIC group was trained on the minimal pairs /tɪnˈdef/ vs. /tɪnˈdeθ/ and /pəˈkif/ vs. /pəˈkiθ/, where the phoneme /f/ had the two different realisations [f] and [ɸ]: they heard /tɪnˈdef/ half of the time as [tɪnˈdef] and half of the time as [tɪnˈdeɸ], and /pəˈkif/ as [pəˈkif] and [pəˈkiɸ]. This group thus learned to treat the non-English fricative [ɸ] as an *allophone*, or more specifically a *free variant*, of the English phoneme /f/. [f] and [ɸ] count as free variants of /f/ because they both occur in the same phonetic context, and there is thus no conditioning factor; if there were a conditioning factor, they would be *conditioned allophones*.

The choice of free variation over conditioned allophony – which was one of convenience – makes no difference to the predictions of the two model types; but it may make a difference to the interpretation of the experimental outcome (as discussed in §12.3).[2]

Finally, a few remarks are in place about my choice of non-native speech sound. Bilabial fricatives are not very common in the world's languages. Voiced bilabial fricatives occur in 54 of the 451 languages in the UPSID database and voiceless bilabial fricatives in 39 (see Maddieson, 1984[3].). But bilabial fricatives are common among languages of the Niger-Kordofanian (or Niger-Congo) family spoken in Western and Southern Africa. In these languages bilabial fricatives often occur as conditioned allophones of /f/; but at least in Ewe, Siya, Logba (all spoken in Ghana), Kwangali (Namibia), Urhobo (Nigeria), Tsonga, Northern Soto and Venda (all South Africa) bilabial fricatives occur as phonemes (see Ladefoged and Maddieson, 1996, pp. 139–143, and Laver, 1994, pp. 253–255). Ewe, Urhobo, Tsonga and Venda are languages that distinguish a voiceless bilabial from voiceless labiodental fricative – the pair used in my training task. An example of a minimal pair from Ewe is /éɸá/ 'he polished' vs. /éfá/ 'he was cold' (where the acute accent indicates a high tone). Outside of Africa, Sinhalese (Sri Lanka), Uzbek and several American and Papuan languages are also reported to have voiceless bilabial fricatives (again according to the UPSID database).

My reason for choosing a bilabial–labiodental fricative contrast despite its relative rarity is twofold. First, I needed a pair where one is an English phoneme and the other a similar sound that English speakers are unlikely to have ever encountered. Secondly, I wanted to have a pair of sounds that can be distinguished auditorily but that can equally easily be treated as belonging to the same phonemic category; more specifically, participants had to accept the new sound as a (maybe slightly unusual) variant of an existing sound of English. There are, of course, plenty of examples from the world's languages of fairly different sounds being treated as realisations of the same phoneme – the English clear vs. dark l [l vs. ɫ] is a case in point. But this kind of allophonic variation is acquired from birth, and it is not clear whether subjects would accept large differences between newly introduced allophones in an experimental context where only little exposure is to make a difference in a subsequent phonetic categorisation task. Bilabial

[2] Note the similarities and differences between my training task and the one use in the perceptual learning paradigm (Norris et al., 2003; see §4.4). In the perceptual learning paradigm, subjects learn to treat an ambiguous fricative sound as either being an /f/ or an /s/. In my training task, subjects in the ALLOPHONIC group also have to treat an unusual sound as belonging to an existing phonetic category (though the training procedure is different); subjects in the PHONEMIC training group, on the other hand, are made to acquire an entirely new category /ɸ/.

[3] Note that Maddieson (1984) was based on an inventory of 317 languages; more were added later. The expanded database with 451 languages is available (at the time of writing) as a free set of MS-DOS applications from http://www.linugistics.ucla.edu/faciliti/sales/software.htm and in the form of a web interface from http://web.phonetik.uni-frankfurt.de/upsid_info.html

fricatives are certainly similar enough to labiodental fricatives to be treated as identical by native speakers of English, but can they also distinguished? Pilot experiments showed that they can, and the outcome of the training task confirmed this (see Chapter 9).

5.2 The repetition priming task

When subjects have acquired the four new words, they performed two tests: a repetition priming and a phonetic categorisation test. The *repetition priming* test consisted of a lexical decision task performed on a list of 360 stimuli some of which were identical or near-identical pairs.

The term *priming*, and even the more specific *repetition priming*, is used to refer to several different experimental paradigms. In Chapter 6, particularly §6.1, I will discuss the different priming paradigms. Repetition priming as used in my experiment has three distinguishing features. In all forms of priming, pairs of stimuli are related as *primes* and *probes*, and we study the influence of the primes on the processing of the probes. Repetition priming is a kind of *form priming*, i.e. primes and probes are related in acoustic, phonetic or phonological form but not in meaning; this is the first distinguishing feature. The second is that subjects have to respond to *all stimuli*, which means that there is no overt distinction between primes and targets. And the third feature is that probes do *not directly follow* their primes: there are stimuli intervening between each prime-probe pair.

In my use of repetition priming, primes and probes were presented all in one block; and participants were asked to perform a lexical decision task, i.e. they had to indicate for each stimulus whether they thought it was a word of English or not. Three types of prime-probe relationship were used: the IDENTICAL, UNRELATED and RELATED conditions.

Priming relationship	**Examples**
IDENTICAL	frɒf–frɒf, bɪˈfɔf–bɪˈfɔf
RELATED	bænf–bænɸ, təˈwɔf–təˈwɔɸ
UNRELATED	brɔf–brɔp, əˈləʊf–əˈləʊt

IDENTICAL pairs were actual repetitions: prime and probe were physically identical. RELATED pairs differed from each other with regard to their final segment only, and the difference involved the non-native [f–ɸ] contrast used in the training. UNRELATED pairs also differed in their final segment; one stimulus ended in [f] and the other in [p,t, or k]. They are called *unrelated* relative to the other two types: they were neither identical nor did they involve the training contrast.

I defined priming as the reaction time to the first-occurring member of a pair minus the reaction time to the second member. A positive value thus means facilitation, and a nega-

tive value inhibition. The RELATED condition is of most interest, since it was the condition in which the training contrast [f–ɸ] occurred, and thus also the condition for which *direct-* and *mediated-access* models make different predictions (see Chapter 8, particularly §8.1). The other two conditions functioned as control conditions and provided reference values: the IDENTICAL condition was taken as the operational definition of full priming, while the UNRELATED condition defined the absence of a priming effect. The UNRELATED condition thus provided a kind of *baseline*, and the IDENTICAL condition a *ceiling*; and between them they delimited the space in which we can assess the results of the RELATED condition.

5.3 The phonetic categorisation task

In the phonetic categorisation test (which was administered immediately after the repetition priming test) subjects performed a categorisation task on two acoustic continua, an OLD and an NEW continuum. Acoustic continua are made in a way that the quality of one of its segments changes from one end of the continuum to the other, from a clear example of one phonetic category to a clear example of another phonetic category. In our case the categories were the labiodental fricative [f] and the bilabial fricative [ɸ].

The following two sets of continua were used:

a) position

OLD	pəˈkif–pəˈkiɸ
NEW	ˈfelə[…]ˈɸelə[…]

b) vowel

OLD	pəˈkif–pəˈkiɸ
NEW	sɑf–sɑɸ

The OLD continuum was derived from the training pair [pəˈkif–pəˈkiɸ], but two different NEW continuum were created. Both were based on pairs which (like the OLD continuum) spanned the [f–ɸ] contrast used in the training, but which (unlike the OLD continuum) subjects had not heard before. In the first NEW continuum, the [f–ɸ] contrast occurred in stimulus-initial instead of stimulus-final position; this continuum is called the *position* continuum, because the position differed from training to test. In the second NEW continuum, the [f–ɸ] contrast occurred in the same position as in the training but in a different vocalic context; this continuum was therefore called the *vowel* continuum. I explain why two NEW continua were used when discussing the predictions of the two types of models (see §8.2).

What subjects have to do in a phonetic categorisation task is say to which category each sound of the continuum belongs. For the OLD continuum, for instance, the question would be whether they hear [pəˈkif] or [pəˈkiɸ].[4] If phonetic categories exist for the sounds which

[4]Because subjects lack labels for the new /ɸ/ category, I have chosen a *categorical AXB* task instead of a direct categorisation or identification task. See §7.5 for further explanation.

form the continuum – the two fricatives [f] and [ɸ], in our case – subjects' performance will be *categorical*, i.e. most sounds of the continuum will be unambiguously assigned to either category. If the categorisation performance is continuous, we may say that the subjects lack the relevant categories. A more formal definition of *categoriality* will follow in §8.2.1.

6 / Methodological review

This second chapter of Part II presents a short review of some issues relevant to the experimental study. The first two sections are relevant to the repetition priming task. I will first discuss different priming paradigms, show what is particular to repetition priming, and explain why I have chosen repetition priming as a test task (§6.1). Then I will look at some determinants of lexical activation and competition (§6.2). And in the last section, I will review phonetic categorisation (§6.3).

One purpose of this methodological review is to acquaint those readers who are not familiar with either repetition priming or phonetic categorisation with the two paradigms. A second purpose is to explain why I chose these two paradigms as test tasks, and to make the details of their implementation (Chapter 7) easier to follow and comprehend. Readers familiar with these topics can go directly to the conclusion (§6.4 on p. 112).

6.1 Priming paradigms in word recognition research

Three types of priming paradigms are commonly used in research on auditory word recognition. The terminology is slightly confusing, as the same paradigms go under different names and the same names are used for different paradigms. I will refer to the three types as *form priming*, *indirect semantic priming* and *repetition priming*.

6.1.1 The three major priming paradigms

In priming as an experimental paradigm we ask how subjects' response to a stimulus – normally in the form of some judgement about that stimulus, or sometimes a simple repetition of the stimulus itself – is influenced by their having heard another, similar stimulus before. Form priming is the most straightforward type of priming and thus may serve as an illustration. In form priming an auditory stimulus, to which subject do not have to respond, is presented immediately before another auditory stimulus, to which subjects are asked to respond. The

two stimuli can be related to each other (e.g. they may contain the same phonemes) or they can be completely unrelated. Performance in related trials is compared with performance in unrelated trials: will subjects' responses be faster and more accurate (facilitation) or slower and less accurate (inhibition) in the related trials? The first stimulus is called the *prime*, and the second the *target* or *probe*.[1]

In general, priming paradigms are about how the processing of a stimulus (the probe) is influenced by the presentation of an earlier stimulus (the prime), where prime and probe are related in some way. The paradigms differ with respect to the following parameters: (i) the type of relationship between prime and probe; (ii) the time interval between prime and probe; and (iii) whether there is an overt distinction between primes and targets.[1]

In *form priming* (Slowiaczek and Pisoni, 1986, Slowiaczek et al., 1987, Radeau et al., 1989, Goldinger et al., 1992, Slowiaczek and Hamburger, 1992, Radeau et al., 1995, Hamburger and Slowiaczek, 1996, Goldinger, 1999, Slowiaczek et al., 2000, Spinelli et al., 2001, Dumay et al., 2001, Gaskell and Marslen-Wilson, 2002, Pitt and Shoaf, 2002, Norris et al., 2002, Dufour and Peereman, 2003, Bölte and Uhe, 2004, McQueen and Sereno, 2005) an overt distinction between primes and targets is made, and subjects only have to respond to the target stimuli. The priming relationship is a phonological one: prime and target have some phonemes in common. Because primes and targets share phonemes, this is generally described as an 'overlap', even though there is of course no physical overlap. Two types of overlap are distinguished: the overlap can be in *stimulus-initial* or *stimulus-final* position. 'Fish' /fɪʃ/ as a prime and 'fog' /fɒg/ as a target have an initial overlap of one segment; 'hat' /hæt/ and 'cat' /kæt/ have a final overlap of two segments (which in this case is also the syllable rhyme). Where there is a partial overlap there also has to be a mismatch; the extent of the mismatch is not normally treated as an extra variable in form priming, and the main explanatory variable is the extent of overlap.

Primes and targets are presented in immediate succession, with an interstimulus interval of between 50 ms and about 1 second.[2] The presentation of stimuli tends to be monomodal (auditory-auditory), but in rare occasions crossmodal presentation has also been used (Dumay et al., 2001, Bölte and Coenen, 2002). Crossmodal presentation allows for the prime and target to be presented at the same time; but it has the drawback that either the prime or the target

[1] For the sake of clarity, I will restrict the term *target* to those cases where each trial contains two stimuli and subjects are asked to respond to the second stimulus only; the stimulus to which they are to respond is the *target*, and the one whose influence on the target we measure the *prime*. When there is no overt distinction, as is the case in repetition priming, there can still be a set of stimuli which serve as *primes* and another set with which we asses the effect of the primes; in this case will use the more general term *probes* for the second set, and not *targets*.

[2] The *interstimulus interval* (or ISI) is measured from the offset of the prime stimulus to the onset of the probe stimulus. Alternatively, it is also measured from prime onset to probe onset; but in this case it is more commonly referred to as a *stimulus onset asynchrony* (SOA).

stimuli have to be presented visually, which makes the concept of a phonological relationship between prime and target somewhat problematic (at least for English).

Indirect semantic priming (Marslen-Wilson and Zwitserlood, 1989, Connine et al., 1993, Marslen-Wilson, 1993, Andruski et al., 1994, Marslen-Wilson et al., 1996, Bölte and Coenen, 2002) can be regarded as a combination of form priming and semantic priming. In *semantic priming*, primes and targets are not related by form but by meaning. Prime and target may be near synonyms (such as 'blank' and 'empty'), belong to the same semantic category (such as 'herring' and 'mackerel'), or may be closely associated (such as 'teacher' and 'pupil'). Semantic priming becomes indirect semantic priming when we start to ask what will happen if we make small changes to the form of the prime, e.g. use 'plank' or 'tank' instead of the semantically related 'blank' as a prime for 'empty'. If the presentation of 'plank' has an influence on the processing of 'empty' this can only happen via the intermediary 'blank'. We must assume that the auditory stimulus /plæŋk/ not only activates the lexical representation PLANK but also BLANK, which in turns activates all its semantically related representations, one of which is EMPTY; and when /ˈɛmptɪ/ is subsequently presented as a target, it will be processed faster because it has already been activated by the prime.[3] Because indirect priming is derived from semantic priming by making small changes to the prime stimulus, it lends itself readily to the study of whether and how mismatching input affects lexical activation. Indirect priming is most often used crossmodally and with simultaneous presentation of prime and target; but intramodal (auditory-auditory) presentation with short ISI has also been used (Marslen-Wilson, 1993, Marslen-Wilson et al., 1996).

The *repetition* or *identity priming* paradigm[4] (e.g. Monsell, 1985, Schacter and Church, 1992, Palmeri et al., 1993, Church and Schacter, 1994, Goldinger, 1996, Luce and Lyons, 1998, Monsell and Hirsh, 1998, Blumstein et al., 2000, Cutler and Donselar, 2001, Pallier et al., 2001, McLennan et al., 2003) is quite different from the two discussed so far. In form priming and indirect semantic priming, every target item has its corresponding prime, the prime immediately precedes the target, and subjects only respond to the target stimulus. In repetition priming, the main distinction is between stimuli which are repeated – either in a completely identical form or with some small deviation – and those which are not. Repetitions are not normally immediately adjacent to each other: primes and probes are often presented in different experimental blocks or session, or – if presented in the same block – there are several stimuli intervening between a prime and its probe.

In general, repeated stimuli are processed faster than non-repeated stimuli. By itself this

[3] Note that the intermediary is itself never presented to the subjects; it is postulated in order to explain the spread of activation from the prime representation to the target representation.

[4] Both terms are sometimes also used to refer to what I am calling *form priming*.

finding may not seem all that interesting; but repetition priming can be used as a test of whether two non-identical stimuli are treated by the perceptual system as similar enough to produce facilitation of a magnitude comparable to straight repetitions. It can thus be regarded as a test to establish the functional identity of non-identical stimuli. By *functional identity* I mean cases where two stimuli are treated as equivalent by the processing system, regardless of whether they are physically identical or not. Repetition priming has been used in this way by the studies discussed in §4.3 above (Pallier et al., 2001 and McLennan et al., 2003, among others; and this is also how it will be used in my experiment.

6.1.2 Form priming and indirect semantic priming

Now that we have seen how priming works and what differences there are between the three paradigms, it is time to consider the result produced by the paradigms and what they may mean for auditory word recognition. This overview is very brief; for additional information the reader is asked to consult the original studies.

Form priming has produced many inconsistent results, particularly with initial overlap. As Hamburger and Slowiaczek (1996) have shown, these inconsistencies are largely due to strategic effects (response strategies or biases),[5] and once these are taken into account, the outcomes are more consistent. The main findings are as follows:

1) Form priming produces strong *strategic effects*. These occur for both initial-overlap priming (Hamburger and Slowiaczek, 1996, Goldinger, 1999, Hamburger and Slowiaczek, 1999, Pitt and Shoaf, 2002, McQueen and Sereno, 2005) and final-overlap priming (Slowiaczek et al., 2000, Norris et al., 2002, McQueen and Sereno, 2005). These strategic effects are facilitative, and are larger the longer the interstimulus interval (ISI) and the higher the proportion of related trials to unrelated trials. But even with a very short ISI and a low relatedness proportion, facilitative strategic effects do not disappear entirely (Goldinger, 1999, Hamburger and Slowiaczek, 1999, Pitt and Shoaf, 2002).

2) *Initial overlap* priming produces an additional non-strategic effect. This effect is facilitative if there is no mismatch later in the stimulus (Spinelli et al., 2001), but becomes inhibitory if the prime mismatches the target in stimulus-final position (Radeau et al., 1989, Goldinger et al., 1992, Radeau et al., 1995, Hamburger and Slowiaczek, 1996, Goldinger, 1999, Hamburger and Slowiaczek, 1999, Spinelli et al., 2001, Dufour and Peere-

[5] Strategic effects occur when subjects employ response strategies to deal with the demands of a task. Because strategic effects are not generally informative about the process studied and it is therefore important to avoid them, I will consider strategic effects in more detail later in §6.1.3.

man, 2003; but compare McQueen and Sereno, 2005, who report non-strategic facilitative effects with initial overlap priming). There is strong evidence that this effect is lexical (Goldinger et al., 1989, Radeau et al., 1995, Dufour and Peereman, 2003).

3) *Final overlap* priming also seems to produce a non-strategic effect. This effect is facilitative (Slowiaczek et al., 1987, Emmorey, 1989, Corina, 1992, Radeau et al., 1995, Slowiaczek et al., 2000, Dumay et al., 2001, Norris et al., 2002). The amount of facilitation increases with the amount of overlap between prime and target (Slowiaczek et al., 1987, Radeau et al., 1995, Dumay et al., 2001, Norris et al., 2002), and rhyme overlap appears to be the strongest predictor (Slowiaczek et al., 2000). It has been suggested that final overlap facilitation has a prelexical locus (Slowiaczek and Hamburger, 1992, Slowiaczek et al., 2000), but, as we have seen in §4.1, the evidence – while consistent with this assumption – is not compelling.

Finding 2 – that initial overlap priming is inhibitory if a mismatch occurs later in the stimulus – and finding 1 – that there are facilitative strategic effects with form priming – can explain why the early results for final overlap priming were inconsistent. If an experimental task is highly susceptible to strategic effects, strategic facilitation may cancel out or even dominate the inhibitory effect of initial overlap; but if we reduce the potential for strategic processing – by reducing ISI and the proportion of related trials – we should find more inhibition. This is what has generally been found (Hamburger and Slowiaczek, 1996, Goldinger, 1999, Dufour and Peereman, 2003). In the case of final overlap priming, on the other hand, both the non-strategic and strategic effects have the same direction: they are both facilitative. This would explain why results have been more consistent with final than with initial overlap priming.

Indirect semantic priming has produced the following results:

1) Priming with semantically related items leads to facilitation: this is the defining characteristics of *semantic* and *associative priming* (Meyer and Schvaneveldt, 1971, Fischler, 1977, Neely, 1977; for recent overviews see Lucas, 2000, Hutchison, 2003).[6] Small changes to the form of the semantically related primes will reduce priming: this is how *indirect priming* works.

2) Changes to the *first segment* of the prime of less than two phonetic features may still result in facilitation, but a change of more than two features makes the facilitation disappear (Marslen-Wilson and Zwitserlood, 1989, Connine et al., 1993, Marslen-Wilson,

[6]Note that most of the studies which have used semantic priming have been carried out with visual and not auditory stimuli.

1993, Marslen-Wilson et al., 1996).[7] Whether facilitation is produced and the size of the effect also depend on when the target is presented relative to the prime and the mode of presentation.

3) Changes to *later segments* produce similar reductions in the facilitative effect of semantic priming (Connine et al., 1993).

4) Whether indirect semantic priming produces *strategic effects* is not known. However, semantic priming can be subject to strategic effects (Neely, 1977, Seidenberg et al., 1984; see also Lucas, 2000). The changes made to the stimuli in indirect semantic priming may conceal the relationship between prime and target; in this case we would not expect strategic effects. Whether this is true for small changes (of two features or less) is not clear. The very short ISI (often 0 ms) may also help to reduce strategic processing.

Despite obvious differences, form priming and semantic priming agree that lexical priming effects are facilitative. A semantically related prime will speed up the processing of a target stimulus, and so does a fragment prime that shares initial phonemes. The mechanisms involved are different, in that the relationship between primes and targets is a semantically mediated one in indirect semantic priming. This explains why even a small change to the prime can make the facilitation disappear, as it can sever the semantic relationship between prime and target. In form priming, we have seen that an initial-overlap prime with subsequent mismatch results in inhibition instead of facilitation. This may be explained as follows. Without mismatch, the prime activates similar lexical representations, among which the lexical representation of the target: this speeds up the processing of the target. With a mismatch, the target representation could be deactivated and one of its closest competitor become highly activated, both of which would slow down the processing of the target.[8]

6.1.3 Strategic effects in priming

One other important finding is that form priming, and presumably also indirect semantic priming, is subject to strategic effects. Strategic effects need to be distinguished from automatic effects (see Goldinger et al., 1992, Hamburger and Slowiaczek, 1996, Goldinger, 1999, Hamburger and Slowiaczek, 1999, Norris et al., 2002, McQueen and Sereno, 2005).

[7] These studies have used the paradigm with crossmodal presentation (auditory primes and visual targets). See Tabossi (1996) for an overview of the paradigm.

[8] There is evidence to suggest that both competition from a highly activated competitor representation (Gaskell and Marslen-Wilson, 2002, Dufour and Peereman, 2003) and deactivation of the target representation (Frauenfelder et al., 2001) may contribute to the inhibitory effect of initial-overlap priming with mismatch.

Automatic effects are effects that are assumed to reflect the process under study – auditory word recognition in the present case. *Strategic effects* are the result of a *response strategy* or *response bias* that subjects develop in order to deal with the demands of an experimental task. Take for example an initial-overlap priming experiment that uses a shadowing task (where subjects are asked to repeat the target stimulus as quickly and accurately as possible). If the proportion of related trials is high, subjects may anticipate the target when they hear the prime; i.e. they expect the initial phonemes of the target to be the same as that of the prime, and prepare themselves to utter these phonemes.[9] This will speed up their responses in related trials and slow them down in unrelated ones, resulting in a large facilitative effect for related trials. This effect, however, is due to a response strategy and does therefore not provide us with any information about word recognition – at least not any unambiguously interpretable information.

Because automatic effects are those which are essential to word recognition, we expect them not to be significantly affected by experimental manipulation that alter subjects' expectations, and neither by other extraneous changes. Strategic effects, on the other hand, are likely to be affected by such changes, particularly those that alter expectations. Two manipulations which are known to affect subjects' expectations are changes to the interstimulus interval (longer ISI afford subjects more time to prepare their responses), and the proportion of related to unrelated trials (a high proportion of related to unrelated trials makes it more likely that subjects notice the relatedness, and it also makes response strategies more beneficial). Changes to the interstimulus interval and the relatedness proportion can thus be used as tests for the presence of strategic effects: if facilitation increases substantially with longer ISI and higher proportion of unrelated trials, the effect is at least in part strategic.[10]

In theory, strategic effects are clearly different from automatic effects, and there are tests for determining whether an effect is strategic or not. But because strategic effects can occur along with automatic effects, they can be hard to identify in practice. Form priming is a good case in point. Since strategic and automatic effects in initial overlap priming with mismatch go in opposite directions, early studies tended to report null results. Circumstances where strategic and automatic effects have the same direction, as is the case with final-overlap priming, will produce fewer ambiguous outcomes. But in these circumstances, it is difficult to be sure whether an effect that has been shown to be strategic does not also has an automatic component.

[9] Subjects need not be aware of the strategy they employ; in such a case it may be preferable to speak of response *biases* instead of *strategies*.

[10] Another way to check for strategic processing is to look at unrelated trials. Strategic processing which results in a facilitation in related trials should come at a cost in unrelated ones: subjects' responses will be inhibited, because their strategy does not work in these cases. See Goldinger (e.g. 1999).

To conclude this short excursion about strategic processing, we can conclude that (i) strategic effects depend on subjects employing some form of response strategy or bias, and (ii) strategic effects can be identified by making changes that affect their usefulness and the likelihood of their occurrence. It is not clear whether priming experiments can be designed in a way that makes strategic effects disappear completely. Current evidence suggests that, at least in form priming, this may not be possible (Goldinger, 1999, Pitt and Shoaf, 2002, Norris et al., 2002).

6.1.4 Repetition priming

As I have described in §6.1, repetition priming is very different from the other two paradigms. In form priming and semantic priming, subjects are presented with prime-target pairs, while repetition priming is not an overt priming task: subjects are just asked to perform a certain task (such as lexical decision) on a list of stimuli, some of which happen to be repetitions or near-repetitions from a previous experimental block or from within the same block. Repetition priming has produced the following findings:

1) The repetition of a stimulus will result in facilitation; this is the basic phenomenon of repetition priming.

2) Small differences between the prime occurrence and the probe occurrence of a stimulus may either produce a similar amount of facilitation as a straight repetition, or they may significantly reduce the amount of facilitation or make it disappear entirely. Differences between primes and probes that have been studied involved:

    - the task, i.e. the same stimuli are presented in different tasks (Monsell, 1985, Schacter and Church, 1992, McLennan et al., 2003);
    - the identity of the speaker, i.e. the same items are spoken by different speakers (Schacter and Church, 1992, Palmeri et al., 1993, Church and Schacter, 1994, Goldinger, 1996, Luce and Lyons, 1998);
    - the exchange of whole segment (Pallier et al., 2001, McLennan et al., 2003);
    - prosodic properties of the stimuli, e.g. stress, pitch, intonation pattern (Church and Schacter, 1994, Cutler and Donselar, 2001);
    - etc.

3) At least in one case, inhibition has been reported (Monsell and Hirsh, 1998); this experiment used prime-probe pairs similar to the ones used in form priming, i.e. with some (initial or final) overlap and a mismatch of several phonemes in length.

Two interpretative problems arise from this brief overview. The first is what perceptual mechanism can explain the basic fact of facilitative repetition priming, the second whether finding 3 (inhibition) can be reconciled with the facilitation found in the other studies.

The mechanism that is responsible for repetition priming is likely to be different from the mechanism that explains the other two types. Form priming and indirect semantic priming appear to reflect the activation of lexical representations (or maybe of sublexical representations in the case of final-overlap priming). It is assumed that by the time the target is presented, the activation caused by the prime stimulus has not yet dissipated and can influence the recognition of the target. Effects of repetition priming – where the distance between prime and probe is at least several seconds, often minutes, and in some cases even days or weeks – cannot be explained by the same mechanism.

Two mechanisms have been proposed. The first suggestion is that repetition priming is the result of long-lasting changes to lexical representations (Morton, 1969, 1979, Monsell, 1985). These changes may take the form of a lowering of the recognition threshold of the lexical representation, or an increase in its resting activation level, or some other strengthening of the lexical representation relative to its competitors. An alternative interpretation claims that repetition priming does not depend on long-lasting changes to lexical representations but has an episodic basis, i.e. it involves the recall of the specific event or episode in which the stimulus has been encountered (Jacoby, 1983).[11] Reports that repetition priming occurs even if the priming task is different from the test task (Monsell, 1985; see also McLennan et al., 2003) are incompatible with repetition priming being episodic in this sense. Monsell and Hirsh's (1998) finding that repetition priming can also result in inhibition through lexical competition also favours the first, lexical account.

This brings us to the second problem: how to combine the inhibition found by Monsell and Hirsh (1998) with the facilitation found in the other studies. If we accept that repetition priming is caused by long-lasting changes to lexical representations, it is easy to see how these differences can come about. If the probe is an exact or near repetition of the prime, we find facilitation because the probe representation has been strengthened by the presentation of the prime stimulus. If probe and prime differ by several phonemes, as was the case in Monsell and Hirsh (1998), then the representation that has been strengthened by the prime stimulus is different from the probe representation and is, in general, one of its close competitors; this will result in inhibition.

[11] Note that *episodic* is also sometimes used to refer to the *lexical-trace* or *exemplar* models described in §1.3.2; this is not the meaning it has here. The term *episodic* as used here goes back to Tulving (1972), who distinguished two distinct memory systems: semantic memory (i.e. knowledge) from episodic memory (i.e. records of specific events).

Finally, we need to consider the issue of strategic processing again. We have seen that form priming and semantic priming are susceptible to strategic processing, because subjects can make use of the prime to anticipate the target and therefore plan their response. This will result in faster responses when their expectations are met, but in slower responses when they are not met. With repetition priming this kind of response strategy cannot develop, since primes and probes are not overtly paired. But are there other possibilities for response strategies to develop?

Response strategies depend on regularities, and one obvious type of regularity is the presentation of primes and probes with a uniform distance, i.e. with the same number of intervening stimuli. In a lexical decision task, for example, if subjects know that the next trial will be a repetition of an earlier stimulus, they may prepare their response based on the response they have given the first time. If regularities such as this one are avoided when designing the experiment, repetition priming should be less susceptible to strategic processing than the other two paradigms, where primes and targets always form overt pairs.[12]

6.2 Factors that affect auditory word recognition

In this section I want to give a brief overview of the factors that influence auditory word recognition, in an attempt to determine which factors need to be taken into account in my own study, and how to best control for them.

Some of the main determinants of the speed and accuracy with which words are recognised in experimental studies are:

1) the *frequency* of the word (Taft and Hambly, 1986, Slowiaczek et al., 1987, Goldinger et al., 1989, Luce et al., 1990, Marslen-Wilson, 1990, Magnuson et al., 2003);

2) the *uniqueness* or *recognition point* of the word (Taft and Hambly, 1986, Gaskell and Marslen-Wilson, 2002, Gaskell and Dumay, 2003);

3) the *number* of its competitors (Goldinger et al., 1989, Zwitserlood, 1989, Luce et al., 1990, Shillcock, 1990, Norris et al., 1995, Vitevitch and Luce, 1998, 1999, Dufour and Peereman, 2003); and

4) the *frequency* of these competitors (Luce et al., 1990, Marslen-Wilson, 1990, Vitevitch and Luce, 1998, 1999).

While all these effects are well established, it is not clear which is the best way to measure or model them. For example, the uniqueness point and number of competitors of a word are

[12] Note that in the literature on repetition priming, strategic processing does not generally seem to be regarded as a major problem.

obviously correlated: all else being equal, the later the uniqueness point of a word the more competitors it will have.

6.2.1 Measures of competitor set size and frequency

With regard to competitor frequency and competitor set size, at least three different measures have been proposed: the mean lexical frequency of the competitor set (Luce et al., 1990), the lexical frequency of the target's strongest competitor (Marslen-Wilson, 1990), and the number of competitors weighted by their frequency (Luce et al., 1990, Vitevitch and Luce, 1998). Bard and Shillcock (1993) have shown that all theses measures are strongly correlated. The reason for this is that competitor sets are highly skewed with regard to lexical frequency: competitor sets tend to have only few high-frequency words (often just one) and a vast majority of low-frequency words. The high-frequency words thus dominate the competition; and as long as they are included in a measure – and all the above measures include them – it will capture most of the competition occurring.

Bard and Shillcock (1993, p. 268) argue that from a theoretical point of view the number of competitors weighted by their frequency is the most appropriate measure, because it subsumes measures that only take account of either competitor frequency or competitor set size. But for practical purposes, we can also conclude that it will not normally matter which measure of lexical competition we use, as long as we use one such measure. The correlation between the measures is not perfect, however, and it is possible to tease apart the effect of frequency and competitor set size by orthogonally varying the two variables. This has been done by Luce et al. (1990, pp. 129ff.), who report a small frequency effect in addition to an effect of set size. This does not, however, invalidate the argument that for more typical stimuli the different measures of competition are highly correlated, and that it therefore is of little concern which one we choose.

6.2.2 Definitions of the competitor set

A similar argument can be applied to the issue of how to define the competitor set. There are two prominent definitions. The *cohort* definition includes only competitors that match from word onset: the word-initial cohort as it is called (Marslen-Wilson and Welsh, 1978, Marslen-Wilson, 1990). This is a dynamic definition of the competitor set, because the set gets smaller as more of the stimulus is presented. The second definition, which we may call the *neighbourhood* definition, compares the overall similarity of lexical representations (Luce, 1986, Luce et al., 1990, Luce and Lyons, 1998). In most applications of the concept, similarity has been operationally defined as an edit distance of one, i.e. the neighbourhood of a given word con-

sists of all words that can be derived from it by a deletion, addition or substitution of a single phoneme. Computed in this way, the neighbourhood definition is static, and it does not take into account the fact that competition starts before all of the stimulus has been heard (Dufour and Peereman, 2003).

Both definitions have empirical support. Taft and Hambly (1986), Zwitserlood (1989), Shillcock (1990), Gaskell and Marslen-Wilson (2002), Dufour and Peereman (2003) and Gaskell and Dumay (2003) have used the *cohort* definition to demonstrate competition effects. Goldinger et al. (1989), Luce et al. (1990), Allopenna et al. (1998), Vitevitch and Luce (1998) and Vitevitch and Luce (1999) have used the *neighbourhood* definition for the same purpose. There also have been some more direct comparisons. Allopenna et al. (1998) have used eye-tracking and a visual world paradigm to demonstrate that objects whose name shares the rhyme but not the onset with the name of the object to be identified will be fixated when that name is played; they thus seem to act as competitors. It is doubtful, however, whether these findings tell us much about normal word recognition. In Allopenna et al.'s visual world task, only four objects were presented at any one time. Such a small set of objects encourage subjects to consider all objects present as potential competitors, and the paradigm may thus greatly overestimate the strength of rhyme competition.

Another direct comparison of the two definition was recently carried out by Newman et al. (2005). They carried out Ganong-type phonetic categorisation experiments[13] and lexical decision experiments. For the phonetic categorisation tasks they used pseudoword continua whose end points differed in the size of their competitor sets. It has been found (Newman et al., 1997) that listeners tend to interpret ambiguous items more often as belonging to the category with more competitors. In the 2005 study, the end points had different competitor sets, but they did not have any competitors with a shared onset; any effect found must therefore be caused by rhyme similarity, and would be evidence for the neighbourhood account. In one of their phonetic categorisation experiments, Newman et al. found a significant bias towards the end point with more competitors.

In the first of their lexical decision experiments, Newman et al. (2005) compared two sets of pseudowords, one of which had more competitors according to the cohort definition and the other according to the neighbourhood definition. They reported a higher accuracy for the set with more cohort competitors. There were, however, no effects on reaction time. In the second lexical decision experiment, two groups of subjects were exposed to two different stimulus sets each. For the one group they matched in terms of neighbourhoods and differed in terms of co-

[13] Ganong (1980) used acoustic continua whose end points differed in their lexicality (word vs. pseudoword) to show that there is a lexical bias, i.e. a bias to interpret a stimulus as a word.

horts, for the second group they differed in terms of neighbourhoods but matched in terms of cohorts. In this experiment, Newman et al. found a significant difference in reaction times only when the two sets differed in terms of neighbourhoods. There were, however, no effects on accuracy. Both lexical decision experiments suggest that the neighbourhood definition captures more of the actual competition than the cohort definition.

Newman et al. (2005) have thus found evidence that the neighbourhood definition of competitor sets may be the more adequate. There are again some reasons to question whether these findings apply to the normal word recognition process. First, the reported effects were not very robust.[14] Secondly, the differences were also not very large (e.g. only a 4% difference in accuracy in the phonetic categorisation experiment), which suggests that the cohort definition also accounts for a substantial part of the actual competition. More importantly still, in order to carry out their experiments, Newman et al. needed competitor sets for which the cohort and neighbourhood definitions produced very different set sizes. Such competitor sets are arguably highly atypical. With typical sets, *cohort* and *neighbourhood* set sizes are similar enough so that the predictions they make are at least qualitatively identical. This as has again been noted by Bard and Shillcock (1993, 245).

To illustrate this point, consider the study by Dufour and Peereman (2003). They used a definition of the competitor set that comprises all the words that have the same length as the target and share its two initial phonemes. With this definition, Dufour and Peereman's *large* competitor sets contained an average of 12.6 competitors, and their *small* sets 3.2 competitors (a ratio of 3.9:1). With the original *cohort* definition (sharing of the first phoneme, regardless of length) the two groups contain 653 and 96 competitors on average (6.8:1). And with a *neighbourhood* definition, the averages are 35 and 20 competitors (1.8:1). While the set sizes vary considerably with the definition used, the relationship between the sets remains unchanged: the larger sets are significantly larger regardless of definition.

It thus seems that for most practical purposes either definition of the competitor set can be used. We may speculate that the reason why both definitions have something to contribute is that each captures an essential part of the activation and competition process. The *cohort* definition is dynamic and is therefore a definition that seems to correspond better to the activation processes, as it takes into account the changing nature of the competitor set. The *neighbourhood* definition, on the other hand, is likely to be a better measure of how similarity is computed in the lexicon. If this hypothesis is correct then a combined definition would be most appropriate: one where the competitor set changes as the speech signal unfolds, but where all competitors

[14] Only one of two phonetic categorisation experiments produced a significant effect; and the two lexical decision experiments showed an effect either only on accuracy or only on reaction time, but never both.

compete to the extent that they are similar with the probe stimulus and with each other.

As with the earlier issue of how lexical competition should be measured, for practical purposes it is of little consequence how we define the competitor set. There are interesting theoretical issues reflected in the different definition, as we have seen; but for our current purpose the most important conclusion is that while lexical competition should be taken into account when designing word recognition experiments, it is of less concern which of the two approaches we use.

6.3 Phonetic categorisation

The second paradigm I have decided to use as a test task is phonetic categorisation. This section introduces the phonetic categorisation paradigm to those not familiar with it. It also explains why I have chosen to use a categorisation or identification task only, and no additional discrimination task.[15]

A typical phonetic categorisation task proceeds as follows. Subjects listen to a set of stimuli that form an acoustic continuum. This continuum goes from a clear case of one speech sound to a clear case of another speech sound via several intermediate, ambiguous sounds. Normally the continuum is made by varying only one phonetic feature or acoustic parameter such as voice onset time (VOT) or place of articulation. If we wanted to study VOT, for example, we could use the two syllables /ba/ (very short or even negative VOT) and /pa/ (long VOT) as the end points of the continuum; for the intermediate stimuli VOT would vary in small steps from the /ba/ value to the /pa/ value.[16] Before listening to the stimuli from this continuum, subjects are given two labels ('BA' and 'PA', or 'P' and 'B' in the example) which they are asked to apply to the stimuli they hear. Their responses can be displayed as a categorisation function. FIGURE 6.1 presents an example of a categorisation function of the /b–p/ voicing contrast in English (taken from Lisker and Abramson, 1970).

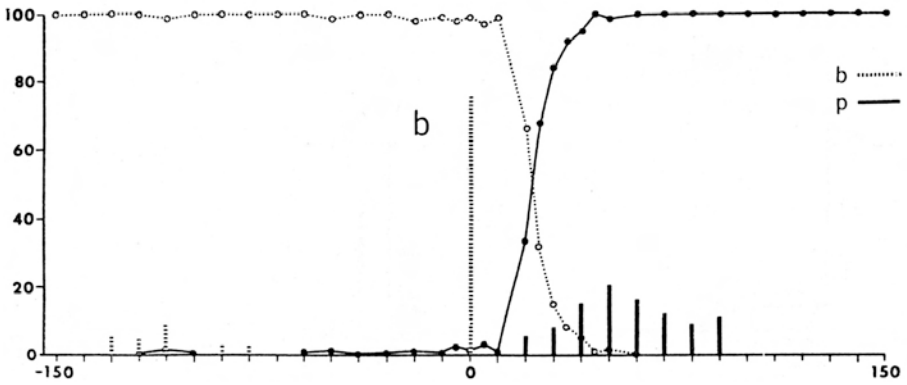
Why can this task be used as a test task in my experiment? Remember that what is needed is a task that can tell us if subjects in the PHONEMIC training group but not those in the ALLOPHONIC group have formed a phonetic category for the new [ɸ] sound. This information could

[15] For this task, both the terms *identification* and *categorisation* have been used in the literature. E.g. Repp (1984) uses *identification* in accordance with the categorical perception literature, and McQueen (1996) uses *categorisation*. In *Detection Theory* it is common to use *identification* when there are as many types of stimuli as there are possible responses; when there are more types of stimuli than possible responses, the term *categorisation* is used (Macmillan and Creelman, 2005, p. 113). If we want to follow this usage, then it would be better to speak of *categorisation* instead of *identification*, because there are commonly only two (or maybe three) responses and at the very least half a dozen of distinct stimuli.

[16] Voice onset time is the temporal difference between the release of a stop consonant and the onset of voicing of the following vowel; this is why stops have to be presented with following vowel (or sonorant). In general speech sounds are presented embedded in syllables or even polysyllabic stimuli.

be provided by categorical perception. *Categorical perception* in its most general sense is the phenomenon that a continuous series of stimuli (such the acoustic continuum of the phonetic categorisation task) is perceived as discontinuous when it crosses the boundary between two categories (Repp, 1984, p. 252f).

a) categorisation function



b) discrimination function

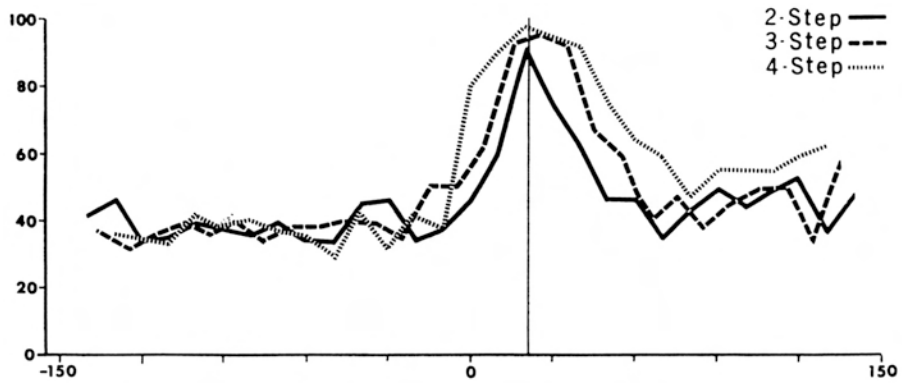


FIGURE 6.1: Categorical perception. Categorisation function and discrimination function of English subjects for a synthetic [b–p] continuum. Taken from Lisker and Abramson (1970) and Abramson and Lisker (1970). The numbers of subjects were 12 (categorisation) and 5 (discrimination).

6.3.1 Categorical perception

The original and best-known operational definition of categorical perception was developed by the Haskins Laboratories and is in terms of a categorisation and discrimination task. There are several different discrimination tasks (see e.g. Macmillan et al., 1977, Macmillan and Creelman, 2005, ch. 9). All of them use pairs of sounds from the continuum, with a fixed distance or step-size between them; they differ with regard to how the discrimination is performed. In

the simplest task, the *same-different task*, subjects are presented with stimuli that contain two signals;[17] the signals are either identical or different, and subjects have to indicate whether they hear them as identical or different. In an *ABX* or *AXB task*, subjects are presented with three signals, two of which (A and B) are again the pair taken from the continuum and the third (X) is identical with either A or B; subjects are asked to indicate whether X is identical to A or B. In an *oddity task* subjects are given stimuli with three signals, one of which is different from the other two, as in an AXB task; but now they have to indicate which signal is the odd one out.

Discrimination tasks result in a discrimination function (see FIGURE 6.1, bottom), and the Haskins definition of categorical perception relates the discrimination function to the categorisation function: perception is categorical if the discrimination performance is close to perfect across the category boundary as defined by the identification function, and at chance level within a category (Liberman et al., 1957; see also Repp, 1984, pp. 251–254). In other words, discrimination is constrained by the categories available in the sense that subjects cannot discriminate within the bounds of a category. In the example in FIGURE 6.1 (taken from Lisker and Abramson, 1970, Abramson and Lisker, 1970) this relationship between categorisation and discrimination indeed holds. But notice that the discrimination performance with the 2-step continuum only reaches about 90% across the category boundary; and the 4-step continuum, which comes closer to eliciting an ideal performance across the boundary, produces a within-category performance for /p/ that is somewhat better than chance.

The categorical perception paradigm could be used as a test in my experiment in the following way. We let subjects perform a categorisation and a discrimination task on an [f–ɸ] continuum. The prediction (at least of the *mediated-access* model) would be that the PHONEMIC group would show categorical perception and the ALLOPHONIC group would not, because only the former have learnt to treat regard [ɸ] as a category distinct from [f]. This would mean that only for the PHONEMIC training group would the discrimination performance be constrained by their categorisation performance.

The Haskins definition of categorical perception is attractive because it is clear and operational. It does have some major drawbacks however. The first is that the definition is not specific enough. Discrimination performance depends heavily on task factors (Repp, 1984, pp. 259–272 for an overview). Tasks differ in their sensitivity: with more sensitive tasks performance tends to be fairly continuous (Schouten et al., 2003, Gerrits and Schouten, 2004), suggesting that categories put no absolute constraints on discrimination performance. This is a problem for the Haskins definition; but one that could in principle be solved by being more

[17] It is common to refer to the whole (pair, triplet, etc.) as the *stimulus* and the individual sounds that subjects have to compared as *signals*.

specific about which task has to be used, and with what interstimulus interval, etc.

The second problem is that not all speech categories are perceived equally categorically. Vowels, for example, tend to have a more continuous discrimination function than stop consonants (Fry et al., 1962, Stevens et al., 1969, Pisoni, 1973, 1975).[18] But even for stop consonants – which appear to be the most categorically perceived speech sounds – it seems that discrimination performance is less categorical when stops are presented in syllable-final position than in syllable-initial position (Rapahel, 1972, Miller et al., 1979). That not all categories produce categorical perception as defined in the Haskins way, is very problematic. If for a large number of categories, categorisation performance is not predictive of discrimination performance, than the Haskins definition of categorical perception is not appropriate as a test for the existence of categories.

In response to this second problem, we might say that we do not need to worry about categorical perception not being a universally applicable test for the existence of categories, as long as it is a test that works for the kinds of categories we are looking at, namely fricatives. Fricatives have produced conflicting results (Repp, 1984, p. 286f). Some studies found fairly categorical discrimination performances (May, 1981, Repp, 1981b), but within-category discrimination has often been better than chance (Fujisaki and Kawashima, 1969, 1970, Healy and Repp, 1982). We thus cannot know whether the Haskins definition would work in the case of [f–ɸ] continua.

Finally, it has also been claimed that discrimination performance may be determined by psychophysical factors; and that rather than discontinuities in discrimination performance being a consequence of category boundaries, languages may make use of existing psychophysical discontinuities when 'choosing' their categories (Pastore et al., 1977, Stevens, 1981, Pastore, 1987, Stevens, 1989). If this is true, the Haskins definition of categorical perception would not only fail for certain types of categories, it would be completely inadequate.

6.3.2 Phonetic categorisation without discrimination

We do not need to go as far as to entirely reject the Haskins definition of categorical perception, even if there are reasons for doing so (see e.g. Macmillan, 1987, Massaro, 1987); but it is clear that it is not a good way of determining whether a new category has been formed by the subjects of my two training groups. A common practice is to only use a phonetic categorisation task and to compare the categorisation functions of different continua, the same continuum in different

[18] The situation is complicated by the fact there also are differences in how sounds are affected by changes in task factors. Pisoni (1973) has shown that vowel discrimination becomes less categorical with longer ISI, while stop consonant discrimination remains relatively stable.

contexts, or the same continuum with different populations. Different continua have been used in trading relations (e.g. Summerfield and Haggard, 1977, Fitch et al., 1980) and lexical bias experiments (e.g. Ganong, 1980, Fox, 1984, McQueen, 1991); different contexts have been used in selective adaptation experiments (e.g. Eimas and Corbit, 1973, Sawusch and Jusczyk, 1981); and the use of different populations included comparisons of adults and children (e.g. Nittrouer and Studdert-Kennedy, 1987, Mayo and Turk, 2004), native speakers and learners (e.g. Flege and Hillenbrand, 1986, Flege, 1992), humans and animals (Kuhl and Miller, 1978), but also of groups that have undergone different training regimes (e.g. Pisoni et al., 1982, Norris et al., 2003). This is the case in my experiment as well.

What researchers tend to look for in these cases are shifts in the location of the category boundary. This can be best exemplified with a trading relations experiment. Fitch et al. (1980) constructed synthetic [slɪt] and [splɪt] continua by varying the duration of the silent interval after the fricative. If the interval exceeded a certain duration – thus indicating a stop closure – subjects would interpret the stimuli as 'split'. For the continuum based on [splɪt], which had vowel transitions appropriate for a /p/, the category boundary between 'slit' and 'split' responses occurred at a lower silent duration than for the continuum based on [slɪt], which lacked vowel transitions that cued /p/.

Norris et al. (2003), with their perceptual learning paradigm, have shown that making two groups acquire different lexical biases, results in different category boundaries on an [ɛf–ɛs] continuum (as discussed in detail in §4.4). My experiment is similar to this paradigm, as it also involves training; but it differs in that the training is meant to make one of the group acquire a new phonemic category rather than to induce a lexical bias. What I am looking for in the phonetic categorisation task is, consequently, not a shift in the location of the category boundary, but a between-group difference in the *degree of categoriality* of the responses. How I measured differences in categoriality is described in §8.2, and problems in interpreting categorisation performance in §12.3.4.

6.4 Conclusions

I have chosen repetition priming as my main test task, because it is the only task that has been used explicitly as a test for what I have chosen to call *functional identity*, i.e. as a test for whether two physically different stimuli are treated as the same by human listeners (see e.g. Pallier et al., 2001, McLennan et al., 2003). Form priming would be difficult to adapt for this purpose, partly because of the conflicting results, but mainly because even small overlap between primes and targets will produce an effect. An additional drawback is that form priming has been shown to

be very susceptible to strategic processing. Indirect semantic priming could be used as a test for *functional identity* – a prime which causes as much facilitation as the actual semantically related prime can be regarded as functionally identical to it – but, to my knowledge, it has never been used in that way. It would also be slightly more difficult to implement, because of the need to have semantically related prime-probe pairs.

My brief discussion of strategic processing has shown that repetition priming does not encourage strategic processing, as long as we avoid regularities in the construction of the test lists which could be used to anticipate the lexical status of stimuli (I will be using a lexical decision task). How I have tried to avoid such regularities is explained in §7.4.

The discussion of factors that influence word recognition has shown that we should take lexical competition into account, but that it does matter less how we define the competitor set and how we measure competition. Because my test stimuli are all pseudowords that diverge on the last segment from real words (§7.4.1 spells out the reason for this), I have chosen a cohort definition, and will measure competition by the lexical frequency of the closest competitor. This seemed to be the most natural way of controlling for competition, given the way test stimuli have been selected and constructed.

Finally, in the section on phonetic categorisation, I hope to have shown why I consider phonetic categorisation without a discrimination task the most appropriate way of testing for the existence of phonological categories for the speech sound [ɸ] introduced in the training. My choice of task (*categorical AXB*), and how I will compare the training groups with regard to the degree of categoriality of their performance, is explained in §7.5 and §8.2.

7 / Method

This chapter describes how I have implemented the experimental design. I first describe the recruitment of participants (§7.1), followed by the equipment (§7.2), then present the training task (§7.3) and the two test tasks (repetition priming in §7.4 and phonetic categorisation in §7.5). A brief comment on the statistical methods used (§7.6) will complete the chapter.

7.1 Participants

A total of 68 participants took part in the study (not including several pilot experiments). They were recruited through an advertisement on the University of Edinburgh's Careers Service web page, and were all students at the University of Edinburgh. Participants were paid a total of 15 pounds for taking part in the whole study: £3 each for the two short training sessions, £5 for the longer test session, and the remaining £4 as a bonus for finishing the experiment.

The study was run in three separate series of experiments, each consisting of two training sessions and one test session. Because the PHONEMIC training was expected to be harder – as this group had to learn to distinguish minimal pairs spanning a non-native contrast – more participants were initially assigned to the PHONEMIC group in all three series. The aim was to end up with equal numbers of subjects in both training groups.

A total of 33 participants took part in the first training session of Series 1, which was carried out in February 2006; 14 were assigned to the ALLOPHONIC and 19 to the PHONEMIC group. Five subjects of the PHONEMIC group were excluded after the first session because they did not meet the criteria for inclusion (described below in §7.3.2). This means that a total of 28 – 14 in each group – took part in the Series 1. All subjects in this series performed the phonetic categorisation test with the POSITION continua.

Series 2 took place in May 2006. It was begun with a total of 29 participants, 11 in the ALLOPHONIC and 17 in the PHONEMIC group. Five subjects were excluded from the PHONEMIC group, and one subjects in the ALLOPHONIC group failed to turn up for the second training

session. Series 2 was thus completed by 22 participants – 12 in the PHONEMIC and 10 in the ALLOPHONIC group. All subjects that took part in the second series did the phonetic categorisation task with the VOWEL continua.

Series 3 was run because not enough subjects could be recruited for Series 2 (which took place during the exam period). I also tried to achieve equal numbers of subjects in the training groups as well as for the two sets of AXB continua. Series 3 took place in July 2006, and was begun with 10 subjects in the ALLOPHONIC and 11 in the PHONEMIC group. Three subject in the PHONEMIC group had to be excluded after the first training session, resulting in 10 participants in the ALLOPHONIC and 8 in the PHONEMIC group. Three subjects from the PHONEMIC group were tested with the POSITION continua and 5 with the VOWEL continua; in the ALLOPHONIC group the corresponding numbers were 3 and 7, respectively.

The aim of equal numbers in the training groups could thus be met: there were 34 participants in both the ALLOPHONIC and the PHONEMIC training group. Half of each group were tested on the POSITION and half on the VOWEL continuum. All 68 participants were self-declared monolingual native speakers of English, and none of them reported any hearing deficit. Their average age was 21 years and 1 month (sd = 2 years, 1 month); 49 were female and 19 male. Participants spoke various dialects of English; a rough classification revealed 24 speakers of Southern British English, 15 Scottish English speakers, 13 Northern British English, 8 North American English (USA and Canada), and 8 speakers of other varieties (the majority Northern Irish, but also New Zealand and one speaker from Hong Kong). Roughly half the participants (38) knew at least one foreign language. None of the participants was a student of linguistics, and none spoke any languages that have bilabial fricatives.

7.2 Equipment

All experiments were carried out in the speech perception laboratory of the department of Linguistics and English Language at the University of Edinburgh. The laboratory consists of four sound-attenuated and identically equipped booths. The experiments were run with E-Prime, version 1.1.4.1, the stimulus presentation and data collection software of Psychology Software Tools (PST), on Dell Optiplex GX 110 computers. Auditory stimuli were presented binaurally over Sennheiser EH 2270 closed headphones, and visual stimuli on Iiyama TXA 3823 MT monitors. Data were collected either via Dell Quiet Key keyboards or PST's serial response boxes with Audio-Technica ATR 20 dynamic unidirectional microphones for the collection of oral responses.

All auditory stimuli were recorded in the recording studio of the Department of Linguis-

tics and English Language at the University of Edinburgh. Recordings were made directly to hard disk with a sampling rate of 48 kHz (the default). Because E-Prime can only accommodate a limited number of sampling frequencies, the sound files had to be resampled at 44 kHz (44,100 Hz). Also for compatibility all recordings were stored in the WAV sound file format. Editing of sound files was carried out with Syntrillium Software's Cool Edit Pro, version 2.1, and Paul Boersma and David Weenink's Praat, several versions (Boersma and Weenink, 2006).

7.3 The training task

All three series began with two training session. The purpose of the training was that subjects would successfully acquire their four training words. This section describes how the training stimuli were selected and the training procedure that was used.

7.3.1 Materials

a) phonemic training group

b) allophonic training group

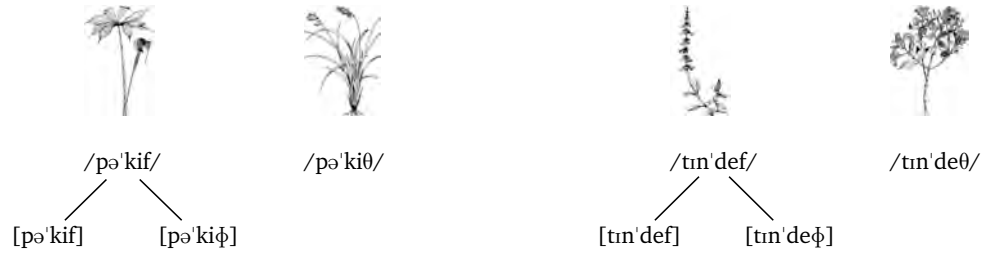
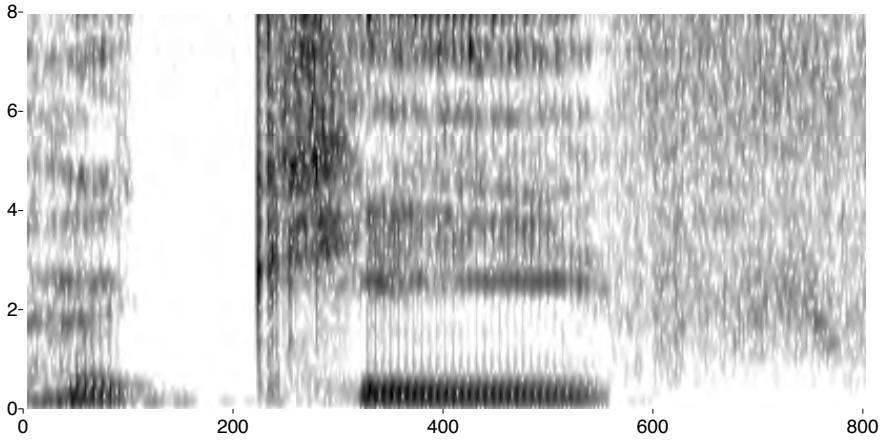


FIGURE 7.1: Minimal pairs used in the training sessions, with the image that specifies their meaning. The PHONEMIC training group will hear the f- and ɸ-stimuli with different images; the ALLOPHONIC group will hear them with the same image.

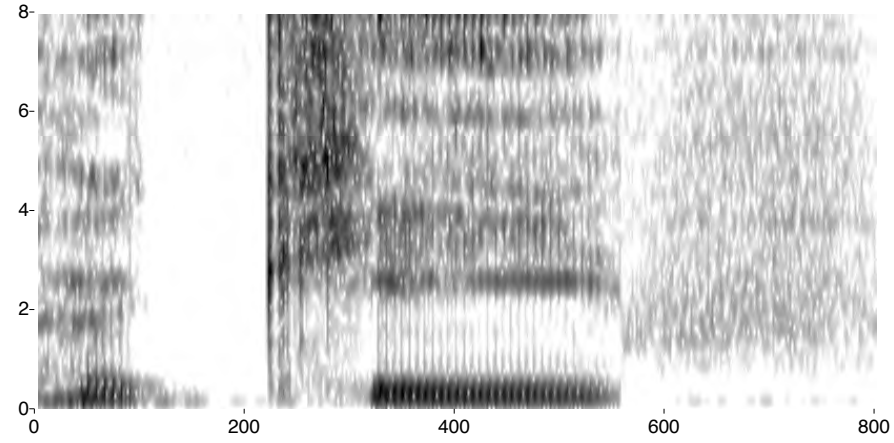
Figure 7.1 (repeated from p. 89) presents the auditory stimuli of both the ALLOPHONIC and PHONEMIC training group, and their meanings, i.e. the images that participants are shown with each auditory stimulus. The pairing of auditory and visual stimuli required the PHONEMIC group to treat [pəˈkif] vs. [pəˈkiɸ], and [tɪnˈdef] vs. [tɪnˈdeɸ] as minimal pairs; the ALLOPHONIC

group was trained to regard them as free variants of the words /pəˈkif/ and /tɪnˈdef/ (see §5.1).

Auditory stimuli. The crucial distinction between the members of each minimal pair occurred in stimulus-final position. This was a requirement of the repetition priming task (the reason is given in §7.4). Apart from the stimulus-final difference, members of each minimal pair had to be identical, or else subjects would have been able use some of the other differences to distinguish between them. This would have defeated the whole point of the training: the acquisition of words that differed only with regard to the [f–ɸ]-contrast.



a) pəˈkif



b) pəˈkiɸ

FIGURE 7.2: Auditory training stimuli. Spectrograms of the training pair [pəˈkif] and [pəˈkiɸ]; time in milliseconds on the x-axis and frequency in kilohertz on the y-axis. Note that the two spectrograms are identical up until the final fricative (from about 570 ms onwards).

Only one version each of [pəˈki...] and [tɪnˈde...] served to create the six auditory training stimuli. FIGURE 7.2 illustrates this for the minimal pair [pəˈkif] vs. [pəˈkiɸ]: both spectrograms are identical except for the final fricative. FIGURE 7.3 shows a spectral slice through the two fricatives. Note that [ɸ] has a lower intensity overall, a steeper decline as we move towards the

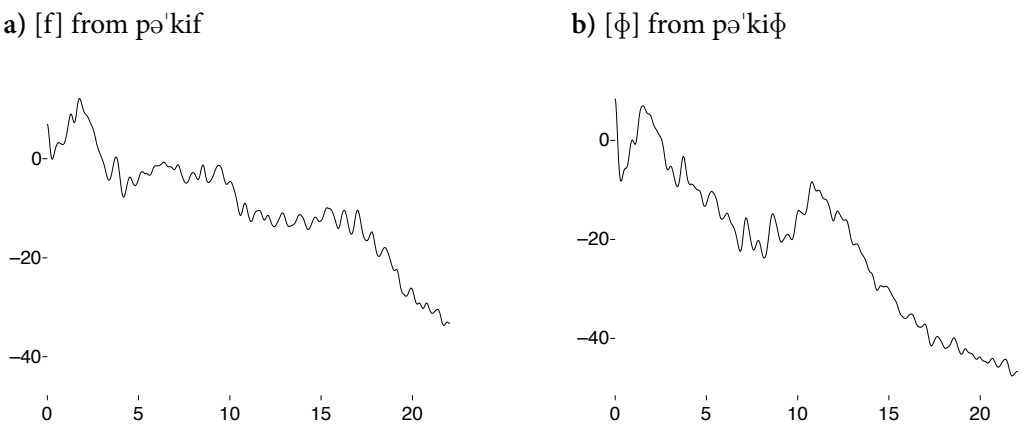


FIGURE 7.3: Auditory training stimuli. Cepstrally smoothed spectral slices through the stimulus-final fricative of the training pair [pəˈkif] and [pəˈkiɸ]; frequency in kilohertz on the x-axis and sound pressure level in dB on the y-axis. Note the different shapes of the two spectra, and the overall lower intensity and steeper decline in intensity of [ɸ] as compared to [f].

higher frequencies, and also a somewhat different shape with a clear second peak around 11 to 12 kHz. The stimulus-final fricatives [f], [ɸ] and [θ] themselves were also the same in both minimal pairs and the allophonic variants.

All auditory stimuli were recorded by the author. Recordings were made in the recording study of the Department of Linguistics and English Language (see §7.2 for details).

Visual stimuli. Because the words used in the repetition priming task were English words, I thought it necessary to introduce the novel words as (uncommon) English words, too. In order to achieve this, the visual stimuli could not be images of objects for which the participants already had a word. I therefore chose four images of plants taken from a monograph on the flora of Sri Lanka (Bond, 1953). Participants were told that the words they were going to learn were used in Sri-Lankan English to refer to local varieties of plants, and that the images are drawings of these plants (see the Consent Form in Appendix C).

The images themselves were chosen to be visually distinctive, so that distinguishing and recognising the images was in itself not hard. As subjects had to learn only four new word, I did not think it necessary to empirically establish visual distinctiveness. Neither participants' performances in the training sessions nor their informal feedback suggested that the distinctiveness of the visual stimuli was a problem. The images were presented to the subjects on a white background with a height of 7.5 cm (corresponding width between 4 and 4.5 cm), and at a viewing distance of 50 to 60 cm.

7.3.2 Procedure

At the start of the the first training session, participants were given an Informed Consent form to complete (see Appendix C). They were informed that they were going to learn four words that were used in a variety of English spoken in Sri Lanka as names of local plants, and that they subsequently had to perform two tests that would assess their acquisition of these new words.[1] Once participants had agreed to take part in the study by completing the form and signing it, they were given a short demonstration of the training procedure. Participants then had the opportunity to ask further questions before they were guided to their individual booth to start the training task. They did not have to perform any practice trials.

Pilot studies run to compare different training procedures had indicated that the acquisition of the novel words was most successful with a sequence of alternating exposure and practice blocks. In the exposure blocks participant simply listened to one auditory stimulus at a time while being presented with the corresponding image. The minimal pairs were presented in immediate succession so as to draw participants' attention to their difference. In the practice blocks participants would see all four pictures at once, would be played an auditory stimulus, and then had to select the image that corresponded to the word they had heard. Finally, they received immediate feedback about the correctness of their choice; and if their response had been incorrect, both the correct auditory stimulus and their incorrect choice were repeated together with the corresponding images.

The PHONEMIC group received 8 practice blocks with 24 trials per block, each preceded by an exposure block of 16 trials. Subjects in this group thus performed a total of 320 trials (8×40). The ALLOPHONIC group had to complete 8 blocks of 40 exposure and 40 practice trials, or a total of 640 trials (8×80). Training sessions lasted about 20 to 25 minutes for the *phonemic* group, and for the *allophonic* group 30 to 35 minutes. If both training groups had received the same number of exposure and practice trials, the ALLOPHONIC group would have heard the crucial stimuli – i.e. [pəˈkif], [pəˈkiɸ], and [tɪnˈdef], [tɪnˈdeɸ] – half as often as the PHONEMIC group. The different numbers were thus chosen to ensure that the overall exposure to the crucial difference was approximately the same for both training groups. But because feedback was given and because the PHONEMIC group generally made more errors than the ALLOPHONIC group (at least initially), it was not possible to make the amount of exposure exactly equal. Feedback was necessary because it was crucial to the success of the training procedure.

As my aim was not to assess the efficacy of the training procedure, but to ensure that subjects could successfully recognise the four training words, participants who did not achieve a

[1] Incidentally, one of the language spoken in Sri Lanka (Sinhalese) does have voiceless bilabial fricatives; and the images were indeed those of local plants.

predetermined success rate in the first session were excluded from the rest of the experiment. Participants could only continue to the second training session if they got more than 80% correct identification, simultaneously for all stimuli, in at least one of the practice blocks. The main reason for excluding subjects who did not reach this level of performance after the first training session was to makes sure that only subjects who had a good chance of learning the four words were included in the study. The purpose of the whole study is not to assess the efficiency of the training procedure, but to compare subjects of equal proficiency in the tests. More important than the value of the cut-off point of 80% was the requirement that subjects had to exceed it simultaneously with all stimuli. Without this requirement, it would have been possible for subjects to concentrate on one stimulus at a time, e.g. by always choosing the same images for all /tɪnˈdef/ and /tɪnˈdeɸ/ stimuli during one block and then the second image during another block.

7.4 The repetition priming task

The purpose of the repetition priming task was to determine whether the two training groups treated stimuli containing the new speech sound [ɸ] differently. The predictions that *direct-* and *mediated-access* models make regarding the repetition priming test task will be described in §8.1.

Stimuli containing the new sound [ɸ] were paired with stimuli containing [f]; this was the RELATED priming condition. In addition there were two baseline conditions: an IDENTICAL condition, where stimuli containing [f] were paired with themselves, and an UNRELATED condition, where they were paired with stimuli containing voiceless plosives, as the following examples illustrate:

Priming relationship	**Examples**
IDENTICAL	frɒf–frɒf, bɪˈfɔf–bɪˈfɔf
RELATED	bænf–bænɸ, təˈwɔf–təˈwɔɸ
UNRELATED	brɔf–brɔp, əˈləʊf–əˈləʊt

The following sections describe the considerations that went into the selection of the test stimuli, and the choice of filler items and the construction of stimulus lists.

7.4.1 Selection of test stimuli

Position of the crucial difference. Why did the crucial difference between stimulus pairs occur in final position? The short answer is that if the crucial difference had occurred in any

other position, it could result in one and the same stimulus having different *deviation points* for the two training groups. The deviation point is to a pseudoword what the uniqueness point is to a word: it is the point where it deviates from all words in the mental lexicon. It is thus the point at which it becomes possible to reject a pseudoword stimulus as a non-word in a lexical decision task.

In the case of stimuli containing the bilabial fricative [ɸ], the position where [ɸ] occurs may determine the deviation point for the PHONEMIC training group, because for them /ɸ/ is a phoneme. For the ALLOPHONIC group – for whom [ɸ] is just a variant of /f/ – the position of [ɸ] should, on the other hand, have no effect on the deviation point. As an example consider the pseudoword [frenk]. It deviates from its lexical neighbours, such as e.g. *friend*, *friendly* and *French*, at the last segment. For subjects in the ALLOPHONIC group the corresponding RELATED stimulus [ɸrenk] would have its deviation point equally on the last segment. For a subject in the PHONEMIC group there is the possibility that the deviation point is on the first segment instead, as there are no words in that subject's mental lexicon that begin with [ɸ].

Such a difference would mean that reaction times for the two groups could not be directly compared, because what looks like a genuine group difference might simply be an artefact of the difference in the location of the deviation points. This might not be a problem, because we are concerned not with absolute reaction times but with reaction time *differences* between the first and second occurrence of the members of a priming pair. But as we saw with the example of [frenk] and [ɸrenk] above, for the ALLOPHONIC training group RELATED pairs would have identical deviation points, whereas for the PHONEMIC training group deviation points of RELATED pairs would differ.

To avoid this shift in the deviation point due to the occurrence of [ɸ], the deviation point and the position of [f] and [ɸ] had to coincide. This could be accomplished most easily by letting the deviation point and the occurrence of the crucial [f–ɸ] difference coincide in final position. In this way it was possible to construct test stimuli straightforwardly from existing English words by replacing their final consonant with [f] and [ɸ] (and [p, t, k] for the UNRELATED condition) in order to form a pseudoword.

The test set. I decided on 20 stimulus pairs per priming condition. This was a compromise between statistical power (which would increase with larger sets) and experimental feasibility (I did not want to overburden my subjects with a test task that was too long). Because I also wanted each of the stimuli to be usable in all three priming conditions, 60 triplets were needed, one each with final [f], final [ɸ] and final [p,t or k]. All 60 triplets are listed in A.1.

Which of the 60 test stimulus types was used in which priming condition was determined

randomly, as was the order in which the pair occurred. For example, one subject might get the stimulus type [frɒC] – where the 'C' stands for the final consonant – in a IDENTICAL priming relationship; this subject would hear [frɒf] twice. For another subject [frɒC] might occur as a RELATED pair; this subject would get [frɒf] followed further down the list by [frɒɸ], or vice versa. Yet another subject might hear [frɒf] and [frɒp]; for this subject [frɒC] would thus be used in the UNRELATED condition.

Every subject within the same training group was tested with a different set of test stimuli, i.e. a different permutation of the 60 test types over the three priming relationships. Across groups, however, subjects were matched: one of the subject in the PHONEMIC group did thus get the same test set as one of the subjects in the ALLOPHONIC group.[2] The reason for the randomisation was to ensure that any difference between *conditions* can indeed be attributed to the factor *priming relationship* and not to any idiosyncrasies of the stimuli chosen: across all subjects such idiosyncrasies are expected to cancel out. The reason for matching subjects across groups was likewise to make sure that any difference between groups was dependent on the factor *training group* alone and not caused by difference between the stimulus sets.

Monosyllables and disyllables. Test stimuli were both mono- and disyllabic. The motivation for using these two types was to make it less likely for subjects to anticipate the end of a stimulus, and consequently the moment at which to press the response button. If all stimuli were either monosyllabic or disyllabic, anticipation might shorten the average response time and reduce its range, with the consequence of also reducing the size of any potential difference between priming conditions or training groups. Because both reaction time and amount of priming may be different for mono- and disyllables, it was made sure that priming relationships had an equal distribution of syllable types. The IDENTICAL, RELATED and UNRELATED condition thus each had 10 monosyllabic and 10 disyllabic stimulus pairs.

At some point I also envisaged the possibility that a comparison of monosyllabic with disyllabic test stimuli may serve as an additional test of the different predictions of the *direct-* and *mediated-access* models. The four training stimuli are all disyllabic. It might be conceivable, although not very likely, that for the PHONEMIC group some generalisation to other disyllabic words is possible even on a *direct-access* account. A *direct-access* model might thus predict a small difference between the PHONEMIC and ALLOPHONIC group as regards their performance in the RELATED condition. But in this case, we expect this difference to disappear, or at least reduce, with monosyllabic stimuli – because no monosyllables were used in the training.

[2] Note that the distribution of the test and filler stimuli over the whole stimulus list was also randomised within the training group and matched across groups; see §7.4.2.

Such a comparison would only be valid if we can exclude other explanations for why such a difference between mono- and disyllabic stimuli should occur. One very important such confound is the competitor environment of the pseudoword: pseudowords in a large competitor set may be rejected quicker than pseudowords in a small competitor set (see §6.2). To make sure that differences in the competitor environment cannot have a confounding effect, mono- and disyllabic stimuli were matched with regard to their competitor environment.

Competitor environments. In addition, the monosyllabic and disyllabic test stimuli were matched with regard to their competitor environment. Test stimuli were chosen by searching the CELEX lexical database of English (version 2.5, Baayen et al., 1995) for words that – in addition to meeting the structural requirements described in the previous sections – had a lexical frequency of over 50 per million in either the 17.9 million words COBUILD corpus (CobMln) or the 1.3 million words subset of spoken data (CobSMln). Pseudowords were then generated from these words by changing the final consonant to [f], [ɸ] and one of [p, t, k] – none of which could be a word. The database was then searched for other potential competitors of the test pseudoword that may be generated from the pseudoword by changing the final consonant, and the competitor with the highest frequency was chosen as a measure of the competitor environment. This way of assessing the competitor environment is based on a Cohort-style definition of the competitor set, and it acknowledges the fact that the strongest competitor dominates the competition process (see §6.2).

Appendix A.1 lists all test stimuli selected in this way. The last column contains the closest competitor word and its lexical frequency per million. Note that this frequency is the average of the CobMln (all words in the COBUILD corpus) and CobSMln (spoken words only) values, so that it may be lower than the 50 occurrences per million specified as the selection criterion.[3] An informal inspection of the sets of mono- and disyllabic test stimuli suggests that their frequency distributions are comparable. Both sets contain one word of very high frequency (*from* and *about*), about half a dozen words with a frequency above 200, another half-dozen with a frequency above 100, and the rest below 100.

The comparability of the two sets was established formally with a two-sample test. Because the frequency distributions of the competitor environments of the the two stimulus sets are non-normal, a distribution-free test was chosen: the Wilcoxon rank sum test, also know as the Mann-Whitney test. Since monosyllabic stimuli in general have more competitors, particularly if we consider that they may also have to compete against polysyllabic candidates. If potential

[3] The subset of spoken words should be a better indicator of speech than the whole corpus; but a larger corpus will be more representative, because the larger the sample the closer will it resemble the population. I thought that taking the average of both values may be a good compromise between both considerations.

polysyllabic competitors are also considered, it is not always obvious which is the strongest competitor of a monosyllabic item. When a polysyllabic competitor had a higher frequency than the closest monosyllabic competitor, both estimates were used in the test, resulting in a low and high estimate of the overall competitor frequency of the monosyllabic test stimuli.

The Wilcoxon rank sum test indicated that the competitor frequencies of the two sets did not differ significantly, neither with the high estimates for the monosyllabic stimuli (W=447.5, p=.98) nor the low estimates (W=498, p=.48). Thus the null hypothesis that both the competitor environment of both mono- and disyllabic stimuli are the same cannot be rejected, and we may conclude that the two sets are comparable.

7.4.2 Filler items

All 60 test pairs (or 120 stimuli) used in an experiment were pseudowords. In order to have a balanced set of words and pseudowords required for a lexical decision task, at least another 120 filler stimuli were required, which all have to be words. In addition, 40 occurrences of the training stimuli were added to the list (which subjects had to regard as words). Their purpose was to ensure that the training had an effect on the repetition priming task. Without their inclusion, subjects might possibly perform the priming task as if they had never taken part in the training, despite the occurrence of stimuli containing the same novel speech sound [ɸ]. The regular occurrence of the training items should make it more likely that the training had an effect in the repetition priming task.

Filler stimuli were selected to ensure that stimulus sets were balanced with regard to their major structural and distributional properties. Half the test stimuli are monosyllabic and half of them disyllabic; so are the fillers. Test stimuli end in [f], [ɸ] and [p, t, k]; so do the fillers, in equal proportions.[4] Finally, some of the occurrences of test stimuli are repetitions, or should at least be perceived as repetitions; an equal amount of repetitions of filler words was therefore included. Balancing these properties across words and pseudoword required an overall set of 360 stimuli: 120 test stimuli and 240 fillers.

Of the 120 test stimuli, 100 ended in either a labiodental or a bilabial fricative, and 20 (of the UNRELATED pairs) in a stop consonant. To match words and pseudowords in this respect as well – and since all 40 training words end in fricatives – 60 of the filler words all ended in [f]. The rest of the fillers had a final stop consonant. The ratio of stimuli with final fricatives to stimuli with final stops was thus 100 to 80 for both words and pseudowords. Some subsequent

[4] There is an exception to this, due to the introduction of phoneme monitoring as an additional task to increase the difference in facilitation between the two control conditions IDENTICAL and UNRELATED. See §7.4.5 for further details.

alteration was required to accommodate the *phoneme monitoring* task. 20 of the words and 20 of the pseudowords were rerecorded with an alveolar fricative /s/ added to the end of the stimulus; this /s/ was the segment that subjects were asked to monitor.

7.4.3 Stimulus lists

As mentioned in §7.4.1, test stimuli were randomised with regard to which *priming relationship* they belong to. The distribution of test and filler stimuli over the 34 different stimulus lists (one for each subject in each training group) was also decided randomly. The placement of test and filler stimuli over the whole list of 360 stimuli was thus randomised and differed within each training group; but there was one subject in the PHONEMIC group and one in the ALLOPHONIC group who received the same list of stimuli.

Primes preceded the corresponding probes by 8, 10, 12, or 14 list positions; in other words, there are 7, 9, 11, or 13 intervening stimuli between each pair. These distances where modelled on the study by Pallier et al. (2001) and, in a set of pilot studies, proved more effective in generating facilitation than did larger distances. The list position of the test stimuli and repeated filler words, as well as the internal ordering of pairs (i.e. which member of a pair is used as the prime and which as the probe) were determined randomly. The rest of the filler stimuli were likewise randomly distributed over the list. Training items are treated differently in that their position were fixed across all stimulus lists. They were spaced fairly regularly over the list, again with distances of 8, 10, 12 and 14 positions between pairs.

The stimulus lists were generated by a Perl script in the following way (see Appendix B.1). First, an empty list was created. Then, the training stimuli was placed at fixed places in the list. Next, from the total set of 60 test pairs, 20 pairs (10 monosyllabic and 10 disyllabic) were chosen at random for each of the three priming relationships. The 20 pairs from the RELATED condition were randomly ordered and distributed over the list, with prime-probe distances of 8, 10, 12 and 14 (5 pairs for each distance). The same was then done with the IDENTICAL and UNRELATED stimuli, in that order. Finally, the repeated words (also with distances of 8, 10, 12 and 14) were placed in the list, followed by all other filler stimuli. The script produced text files with the names of the stimuli in the appropriate order, which could then be read into E-Prime at runtime to determine the order in which stimuli occur in the experimental session.

Within training groups, each subject received a different list, i.e. a different permutation of the 360 stimuli. Across groups, however, subjects were paired up, so that one subject each from the PHONEMIC and the ALLOPHONIC group were tested on the same list. Randomisation of stimuli across priming condition, stimulus position and subjects should ensure that, in the long run, stimulus-specific properties have no effect on reaction times and priming. Matching

of lists across groups, on the other hand, is intended to there is no difference between the two training groups apart from the difference in training conditions.

7.4.4 Recording and editing of the stimuli

Recordings of all the priming stimuli were made in the recording study of the Department of Linguistics and English Language, as described in §7.2. The speaker was the author.

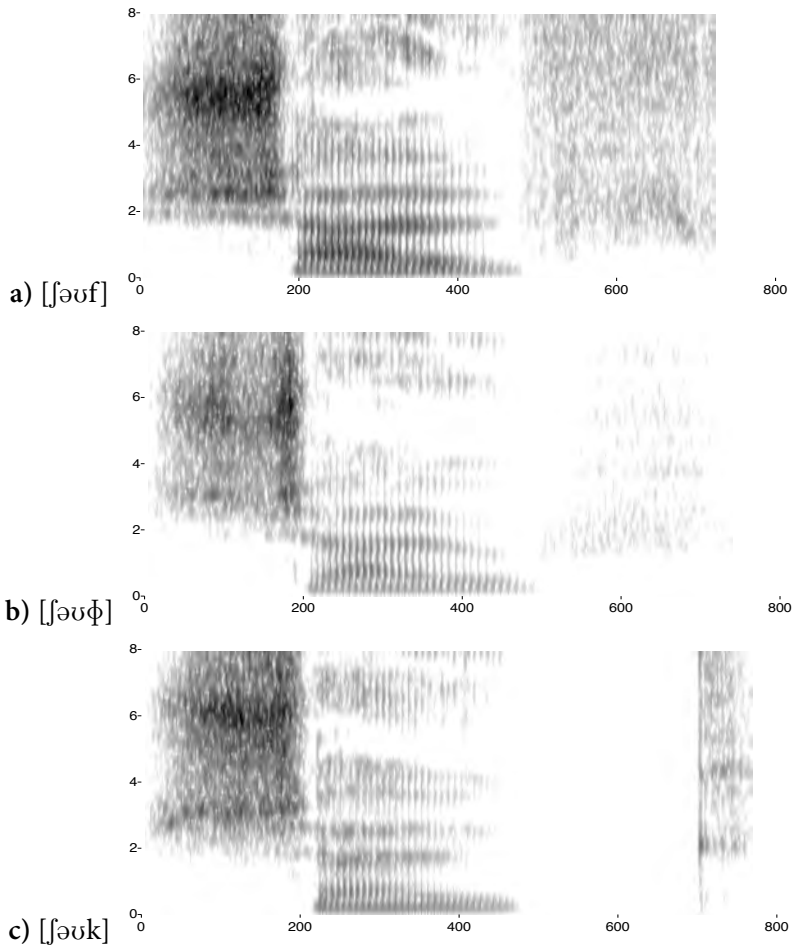


FIGURE 7.4: Priming stimuli. Spectrograms of the priming triplet [ʃəʊf], [ʃəʊɸ] and [ʃəʊk] (bottom). Time in milliseconds on the x-axis, and frequency in kilohertz on the y-axis.

Filler stimuli were excised from the recordings and used without any further editing. The test stimuli were recorded as triplets (one each with final [f], [ɸ] and the stop consonant, as listed in Appendix A.1). In all the stimuli ending in [f] and [ɸ], the final fricative portion was then replaced by the fricatives that were used in the training task; there was thus only one token of [f] and [ɸ] that occurred in all test stimuli. FIGURE 7.4 presents spectrograms of an example

triplet: [ʃəʊf–ʃəʊɸ–ʃəʊk].

One consequence of the way test stimuli were built was that there are some acoustic differences between the priming pairs even before the final consonant, and also differences in their duration; both of this can be seen in FIGURE 7.4. Unlike in the training, where minimal pairs were acoustically identical, apart from the difference in place of articulation of their final fricative, priming pairs were not identical. Priming stimuli were constructed in this way because what is arguably of relevance in word recognition is not acoustic identity but a more abstract, phonological identity: what is being recognised are not word *tokens* but word *types*. Even a *direct-access model* has to acknowledge this – and ought to be able to account for it – because in the actual production of speech no two productions of the same word are absolutely identical. Priming should occur despite these acoustic differences, and that is indeed what previous studies have found.[5]

7.4.5 Procedure

Participants performed a lexical decision task on all 360 stimuli, after having performed a short practice session on 22 stimuli. They were asked to indicate whether they thought they had heard a word or not by pressing buttons on a response box. In addition they had to perform a phoneme monitoring task on the filler items. The reason for introducing this additional task was as follows.

Given what other studies had found (Pallier et al., 2001, McLennan et al., 2003), I was aiming for a difference of about 80 ms between the IDENTICAL and UNRELATED conditions. In the first pilot of the repetition-priming task without any training (and with a sample size of 10 subjects), the difference between the two means was only 33 ms: 91 ms of priming in the IDENTICAL and 58 ms in the UNRELATED condition. For the IDENTICAL condition, the result is of the expected magnitude; however, 58 ms is a rather large value for a condition that is meant to be prototypical for the absence of priming.

I suspected that this fairly large amount of priming even in the UNRELATED condition is an artefact of the stimulus-final position of the crucial difference. In Pallier et al. (2001) and McLennan et al. (2003), for instance, the crucial difference between primes and probes was in stimulus-medial position. It might be that in my experiment subjects made their decision before they had heard the whole stimulus. And since all prime-probe pairs were identical until the very last segment, this behaviour may have resulted in a relatively high amount of priming for all pairs, regardless of priming condition.

[5] It is common practice in repetition priming studies to record primes and probes individually, so that there are naturally occurring differences between them (see e.g. Pallier et al., 2001, McLennan et al., 2003).

To encourage subjects to wait a bit longer before giving their responses, *phoneme monitoring* was introduced as a secondary task. Subjects' task was to spot all stimuli that ended in [s], of which there were 40. Since subjects were instructed to give the phoneme-monitoring response before making the lexical decision, I hoped that phoneme monitoring would induce them to pay more attention to the ending of stimuli, thereby reducing the priming effect in the UNRELATED condition. The introduction of the additional phoneme monitoring task seems to have served its purpose; in a further pilot experiment (with 10 different subjects) the difference between the two control conditions was increased to 83 ms: 116 ms in the IDENTICAL and 33 ms in the UNRELATED condition. The latter is still relatively large, but the difference was of the expected magnitude.

The repetition priming test proceeded in the following way. Participants were first given a demonstration of the lexical decision and phoneme monitoring procedure, before doing a practice set of 22 lexical decisions. They would then do the test task in which they had to perform lexical decisions on the full set of 360 stimuli. This took about 20 minutes to perform. The script could not be stopped by the participants, to ensure that time between pairs was the same for all subjects. They were informed of this, as well as the duration of the task. The stimuli used in the practice block were taken from the set of filler stimuli, except for two training stimuli; most importantly, no test stimuli were used in the practice set.

The procedure used in the repetition-priming task differed slightly between the first experimental series (run in February 2006) and the other two (run in May and July of the same year).[6] The difference was confined to the phoneme-monitoring procedure. The basic lexical-decision task was the same in all three set: stimuli were presented sequentially, and subjects had to decide whether they had heard a word or not by pressing two buttons on a response box. They were also always instructed to treat the training stimuli as words.

In the first series all responses were button presses on a serial response box. Subjects used their dominant hand to perform lexical decisions and the other hand for phoneme monitoring. To proceed to the next auditory stimulus, subjects had to press the middle button (out of five) with the index finger of their dominant hand. They would then let their finger rest on that button until they were ready to make the decision, which was made by pressing the button immediately to the left of the middle one to register a word response, and the button immediately to the right for a nonword response. The phoneme monitoring response was given by pressing either of the outermost button on the response box with their non-dominant hand; right-handed subjects would thus press the far left button, and left-handed subjects the far right button for a monitoring response. Participants were instructed to press this button, if they heard a stimulus

[6] See §7.1 for a description of the three series of experiments.

that ended with the sound /s/, and to give their monitoring response before the lexical decision response. Overall, they had 4 seconds to respond from stimulus onset.

Phoneme-monitoring responses were only required for 40 of the 360 stimuli; most of the time, therefore, participants had to do a straightforward lexical-decision task. Moreover, no monitoring had to be performed on the 120 test stimuli. But the need for subjects to press different buttons with the *same finger* is not an ideal procedure, as the movements between buttons introduces some additional variation. To reduce this potential source of variation, a different monitoring response was introduced in the second and third series of experiments.

In these two series, participants had to give their phoneme-monitoring responses orally: a microphone was attached to each response box, and participants would indicate stimuli with a final /s/ by producing a lingual click sound. This particular response was chosen because the microphones proved very sensitive to click sounds even at relatively low amplitude; click sounds are also easy to produce.[7] Because the monitoring response was now made orally, subjects could use both hands for the lexical decision task. Accordingly, they were instructed to press the far left button with the index finger of their left hand to give a word response, and the far right button with the index finger of their right hand to give an nonword response. Subjects could thus leave their fingers resting on the same button at all time, as there is no need to move between buttons; this was hoped to reduce trial-to-trial variation. Participants were again told to treat training stimuli as words, and to make the phoneme-monitoring response prior to making the lexical decision. Responses had again be given within 4 seconds from stimulus onset.

7.5 The phonetic categorisation task

The phonetic categorisation task also served as a test of the predictions made by *direct-* and *mediated-access* models. Their predictions will be presented in §8.2.

7.5.1 Materials

The continua used in the AXB task were the following:

a) position	OLD	pəˈkif–pəˈkiɸ	**b) vowel**	OLD	pəˈkif–pəˈkiɸ
	NEW	ˈfelət–ˈɸelət		NEW	sɑf–sɑɸ

The OLD continuum was derived from the training pair [pəˈkif–pəˈkiɸ]. The two NEW continua were based on pairs which also span the [f–ɸ] contrast but which subjects had not heard before.

[7] I also considered using pedals for the monitoring responses; but none were readily available.

In the *position* continuum, the [f–ɸ] contrast occurred in stimulus-initial position. In the *vowel* continuum, the [f–ɸ] contrast occurred in the same position as in the training but in a different vocalic context and in a monosyllabic instead of a disyllabic stimulus.

All three continua were constructed in the same way, illustrated here with the OLD continuum. The source of the continuum was the two stimuli [pəˈkif] and [pəˈkiɸ] from the training task. Remember that these stimuli were constructed so as to have the same duration. The method used to produce the continuum was adapted from Repp (1981a). The sounds of the continuum were created as weighted sample-by-sample summations of the amplitude values of the original stimuli. This operation was performed on Praat matrix files of the source sounds with the help Perl script; the script is reproduced in Appendix B.3.

This is how the OLD continuum was produced. A Praat matrix object of a sampled sound is essentially a list of the normalised amplitude values at each sampling point.[8] In order to generate stimuli intermediate between [pəˈkif] and [pəˈkiɸ] on the basis of Praat matrix files, we have to add all the amplitude values in different proportions (e.g. 80% [pəˈkif] and 20% [pəˈkiɸ]]), write them to a new matrix file, and convert this matrix file back to a sound file. A ten-step continuum was produced in this way, i.e. eleven sound files, the first and last of which were identical to the original files, and the other 9 were sample-by-sample combinations of these original files, with the proportions 9:1, 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, 2:8, 1:9. Waveforms of the eleven stimuli produced for the [pəˈkif–pəˈkiɸ] continuum are shown in FIGURE 7.5.

Further illustrations of the categorisation stimuli can be found in FIGURES 7.2 and 7.3 earlier in this chapter. The first shows spectrogrammes of the two training stimuli [pəˈkif] and [p[əˈkiɸ], which also form the endpoints of the OLD continuum illustrated in Figure 7.5. FIGURE 7.3 presents spectra of the two final fricatives [f] and [ɸ]. As can be seen there, the difference between final [f] and [ɸ] is both a difference of overall amplitude – with [f] being higher in amplitude than [ɸ] – and of frequency distribution – with [ɸ] having less of its energy in the higher frequencies. Figure 7.2 also shows that this difference in the final fricative is the only difference between the stimuli.

The eleven sound files were then assembled into triplets for use in the categorisation task. This was an AXB task. The first (A) and last (B) signal of each triplet were always the two end points of the continuum – [pəˈkif] and [pəˈkiɸ] in case of the OLD continuum – and the middle

[8] More specifically "a Matrix object represents a function $z(x, y)$ on the domain $[x_{min}, x_{max}] \times [y_{min}, y_{max}]$. The domain has been sampled in the x and y directions with constant sampling intervals (dx and dy) along each direction. The samples are thus $z[i_y][i_x]$, $i_x = 1 \ldots n_x$, $i_y = 1 \ldots n_y$. The samples represent the function values $z(x_1 + (ix - 1)dx, y_1 + (iy - 1)dy)$. [...] If the matrix represents a sampled signal of 1 second duration with a sampling frequency of 10 kHz, it has the following attributes: $x_{min} = 0.0$; $x_{max} = 1.0$; $n_x = 10000$; $dx = 1.0 \times 10^{-4}$; $x_1 = 0.5 \times 10^{-4}$; $y_{min} = 1$; $y_{max} = 1$; $n_y = 1$; $dy = 1$; $y_1 = 1$." (Boersma and Weenink, 2006) We may say, somewhat simplistically, that x stands for the individual samples and y represents the amplitude value of sample x.

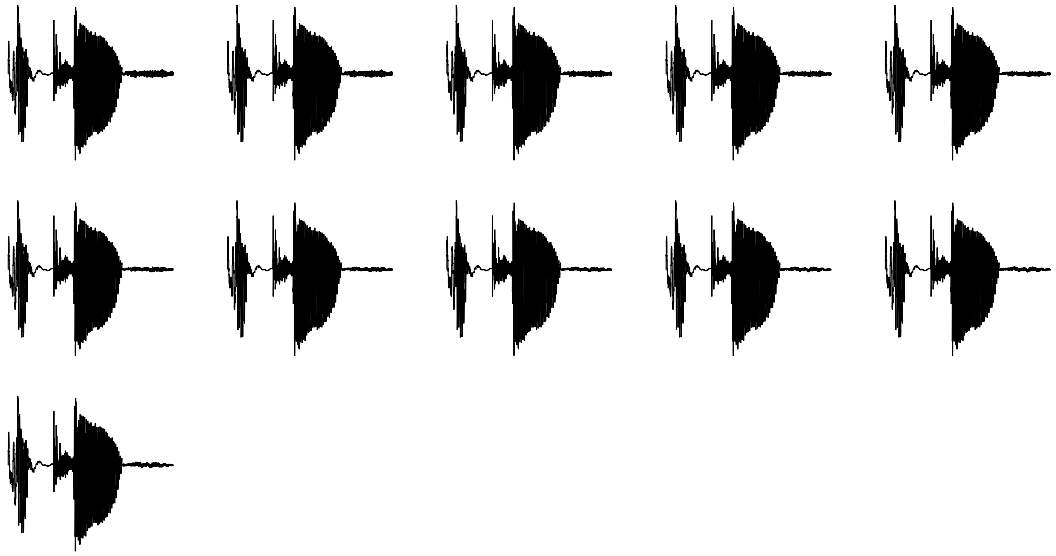


FIGURE 7.5: The OLD continuum. Waveforms corresponding to the 11 sounds of the pəˈkif–pəˈkiɸ continuum. Note how the final fricative portion of the signal changes from the top left (original [pəˈkif] stimulus) to the bottom right (original [pəˈkiɸ] stimulus). The earlier parts of the stimuli are identical.

signal was any of the 11 sounds of the continuum (including the end points). 500 ms of silence separated the A from the X and the X from the B. Because both end points could occur as the first or the last signal – AXB as well as BXA triplets were used, in other words – 22 triplets were generated for each continuum.

The NEW continua were produced in the same way. First the two end points had to be recorded and edited, i.e. [ˈfelət], [ˈɸelət] and [sɑf], [sɑɸ]. As for the OLD continuum, each member of a pair had the same duration (i.e. the same number of samples) and was identical apart from the fricative segment. The two NEW continua and the corresponding stimuli were produced as described for the OLD continuum.

7.5.2 Procedure

The standard categorisation or identification task has been described in §6.3. Subjects are given a set of phonetic labels (generally just two), and are asked to categorise the sounds of an acoustic continuum with these labels. With a [f–ɸ] continuum, where one end is a non-native speech sound, a direct categorisation of stimuli was not possible. A *categorical AXB* or *AXB identification* task can be used in cases like this (see e.g. Best et al., 1981, Polka and Bohn, 1996).

We have already encountered the AXB task as a discrimination task: the A and B are the two sounds to be discriminated, and the X is identical with either A or B; subjects then have to indicate whether the middle signal of each stimulus is identical with the first or the last signal. If

the AXB task is used as a categorisation task, A and B are the two *end points* of the continuum, and X can be any sound along the continuum (including the end points). Subjects have to judge whether X is more similar to A or to B. They are thus in fact performing a similarity judgement, but because the comparison is made with the end points of the continuum, their responses can be interpreted as a categorisation of X as belonging to either A or B.

Each subject made a total of 88 judgements per continuum. As described in the previous section, there were 22 different stimuli for each continuum: one for the 11 different sounds of the continuum presented in the order AXB (i.e. [f]-endpoint first) and BXA (i.e. [ɸ]-endpoint first). All 22 stimuli were presented 4 times in random order ($4 \times 22 = 88$). Since each subjects was tested on both the OLD and one of the NEW continuum, each subjects did a total of 176 categorisation trials (2×88). The trials were presented as 4 blocks of 44 trial each. The OLD continuum was presented first, in blocks 1 and 2, followed by the NEW continuum in blocks 3 and 4.

In each trial, an AXB triplet would be presented to the subject over headphones. They then had to decide whether the middle sound was more like the first or last sound, by pressing the leftmost (for 'first') or rightmost (for 'last') button on the five-button response box. As with the lexical decision task used in the repetition priming test, subjects were instructed to respond as fast and accurately as possible. But unlike in the former task, they were given plenty of time to respond (up to 10 seconds from stimulus onset). The reason for this was that our main response variable in the categorisation task is the actual categorisation response and not the reaction time; and a more leisurely pace, which gives subjects time to make their judgements, should reduce erroneous responses.[9] Subjects could choose for themselves how long they wanted to rest between blocks; resting times were generally short, never more than about a minute.

7.6 Statistical analyses

I have decided to use mixed-effects modelling in order to deal with the repeated observations (on subjects and stimuli) in my data. In this section I will briefly outline why. The use of mixed-effects models over more conventional techniques, particularly the use of by-subject and by-item ANOVAs, has the following advantages.[10]

[9] Note that we cannot tell from our data whether a response was given in error or not. But an erroneous response would be one which the subject would have corrected, had they had the opportunity to do so. And it seems likely that subjects make more such errors when they have to perform under pressure.

[10] A concise introduction to mixed-effects models is provided by Faraway (2006, ch. 8); Maxwell and Delaney (2004, chs. 15 & 16) is a less technical introduction; Baayen et al. (submitted) discusses the use of mixed-effects modelling in psycholinguistics; and Snijders and Bosker (1999) is a book-length introduction from a multilevel perspective.

Mixed-effects models address the problems identified by Clark (1973). As researchers who study languages experimentally, we want to make assertions about a specific language or even about languages in general. In our experiments however, we only look at a sample of speakers and a sample of linguistic materials. If we want to generalise to the whole population of speakers and to all possible linguistic materials of the same type, we have to regard both the subjects and the language materials as a random factors. The failure to do the second has been called the 'language-as-fixed-effects fallacy' in Clark's seminal (1973) article.

The solution that Clark proposed was to use quasi F-ratios *(F')*, or if this is not possible, to compute *minF'* (the minimum value of *F')* as an approximation of *F'*. This advice was optimal at the time, but statistics has moved on since then and techniques such as mixed-effects models have become available and computationally feasible. Mixed-effects models allow the inclusion of both fixed and random effects in the same statistical model. Through this possibility to include random factors for both subjects *and* items in the model, mixed-effects models can provide a general solution to the 'language-as-fixed-effects fallacy'.

Mixed-effects models are a better solution than the alternatives. In spite of Clark's recommendations, the most common method currently in use is to compute separate by-subject and by-item analyses. This procedure is clearly not optimal. If the two analyses disagree we do not know what to conclude. But even if the by-subject and by-item analyses are both significant, we cannot be sure that a joint analysis would also come out significant, as shown by Clark (1973, pp. 341–347).

Clark's solution – quasi-F tests – is better than separate by-items and by-subject analyses. Its drawbacks are that each experimental design requires its own analytic solution, and *minF'* as an estimate of *F'* can be a too conservative.[11] Mixed-effects modelling has the advantage that because it relies on maximum-likelihood estimation it is much more widely applicable than the quasi-F test proposed by Clark.[12]

Raaijmakers et al. (1999) have suggested that in cases where the stimulus sets are carefully matched and counterbalanced, there is no need to include items as an additional random factor and that, therefore, analyses with subjects as the only random factor can be performed. While Raaijmakers et al. may have a case, their argument only applies to some experimental designs (as they themselves point out) and cannot be used as a universal analysis strategy. Mixed-effects models have the advantage that they are more general. And as Baayen (2004) and Baayen et al.

[11] Though Raaijmakers et al. (1999) claim, referring to simulation studies by Davenport and Dickinson (1973) and Forster and Dickinson (1976), that both *F'* and *minF'* are not unduly conservative.

[12] Although it has to be said that maximum-likelihood estimation, because it is non-analytic, depends on numerical estimation algorithms, which have only recently become generally available and are still being developed (see e.g. Pinheiro and Bates, 2000, Bates and Sarkar, 2007).

(submitted) have shown with simulation studies, even in a case where a simple subject analysis could be regarded as appropriate, a mixed-effect analysis has the higher statistical power. More generally, the simulations by Baayen (2004) and Baayen et al. (submitted) suggest that mixed-effects models perform at least as good as the alternatives, both in terms of maintaining the nominal α-level and statistical power.

Mixed-effects models can accommodate unbalanced data. An important advantage of mixed-effects models is that they are robust with respect to missing data. This is a consequence of the use of maximum-likelihood estimation instead of ANOVA or least-squares estimation (see Faraway, 2006, pp.154–158). This was probably the most important reason for choosing mixed-effects models, as there were missing responses in my experiment.

Finally, I will briefly describe how mixed-effects models are used, and how I will apply them to my data. For a comprehensive description mixed-effects modelling see Faraway (2006, ch. 8) and the literature cited earlier.

Corresponding to the F-test used to test for main effects and interactions with ANOVAs, there is the likelihood ratio test with mixed-effects models. This is performed by fitting two models to the data, one with and one without the factor we want to test for, and then perform a likelihood ratio test to see whether the addition of the factor increases the fit of the model significantly. The likelihood ratio test has an approximate χ^2 distribution. But because the χ^2 distribution is only an approximation, it is recommended to use parametric bootstrap estimation when the outcome is equivocal, so as to obtain more accurate *p*-values (Faraway, 2006, pp. 158–161). For pairwise comparisons of levels within a factor a *t*-test could be performed; but this is not advisable, as explained by Baayen et al. (submitted). I have followed their advice and use confidence intervals based on Markov chain Monte Carlo (MCMC) sampling instead.

Both the training task (see Chapter 9) and the phonetic categorisation task (see Chapter 11) produced non-normal responses, in the form of proportions of correct identifications (training task) and proportions of /f/-responses (phonetic categorisation task). The appropriate model in this case is a generalised linear model with a logistic (also called logit) link function and a binomial error distribution – a logistic model in other words.[13] The models I used were mixed-effects variants of the corresponding generalised linear models.

The data analysis and model fitting was done using the statistical software R (R Development

[13] Instead of the logit link function, a probit or a complementary log-log link could have been used. The difference between these choices is, in general, negligible if we remain within the range of the data; differences only appear when we extrapolate to values beyond the data set (Faraway, 2006, 36–38). I have chosen a logit link function because it is mathematically simpler and easier to interpret than the other two. The logit link function is $\eta = log(\mu/(1-\mu))$, where η represents the response variable in the linear model, and μ the actual mean response.

Core Team, 2006) and its mixed-effects modelling package `lme4` (Bates and Sarkar, 2007). The random factors included were *subjects* and *items*, except for the training task. In the training, the different items constituted the two training groups and were thus not included as an additional random factor in the model. Details of the analyses are given in the relevant results chapters (Chapters 9 to 11).

8/ Predictions

In this chapter, I present what I think are the predictions that *direct-* and *mediated-access* models make regarding the test tasks. The experimental design has been described in Chapter 5 and the individual tasks in Chapter 7. Note that there are no predictions about the training task, as it is an integral part of the overall design. The purpose of the training is to generate the difference between training groups on which the predictions for the two test tasks depend. This is why there are no separate predictions for the training task. The data of the training task will nonetheless be analysed in Chapter 9, in order to see whether the training has been successful.

8.1 The repetition priming task

Mediated-access models assume that if listeners have acquired new words that contain the non-native phoneme /ɸ/, they necessarily have developed segmental representations for the non-native phoneme. When the same listeners then process other words that contain the new sound, we can expect them to use these newly developed prelexical representations. *Direct-access* models assume that no prelexical representations exist, and that words are acquired whole. Listeners will therefore not have developed segmental representations for the non-native phoneme /ɸ/. Figure 8.1 shows what, given these fundamental assumptions, the two types of models predict with regard to the repetition priming task.

Mediated-access models predict that participants in the PHONEMIC training group will treat stimulus pairs in the RELATED priming condition – which are identical except that one has [ɸ] where the other has [f] – as different, because they have separate segmental representations for the two sounds; participants in the ALLOPHONIC training group should treat them as identical, because they only have one segmental representation for both [f] and [ɸ]. This means that the PHONEMIC group should produce as little priming in the RELATED condition as they do in the UNRELATED condition, and the ALLOPHONIC group as much priming as in the IDENTICAL condition. These predictions are illustrated on the left-hand side of Figure 8.1: the two broken

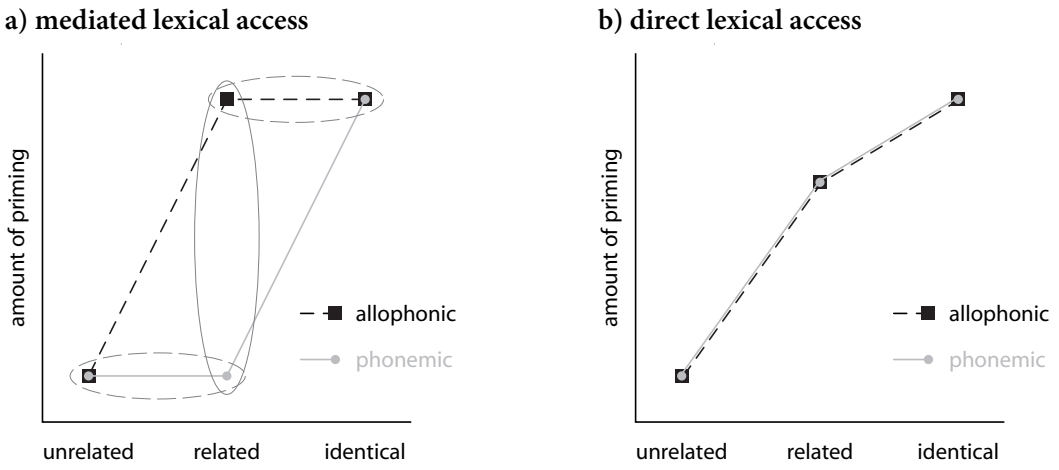


FIGURE 8.1: Repetition priming task: idealised predictions of *mediated-access* models (left) and *direct-access* models (right). See running text for further explanation.

ellipses highlight the conditions where the two training groups are predicted to produce identical amounts of priming; the solid ellipse highlights the resulting between-group difference.

The predictions of *direct-access* models are shown on the right-hand side of the same figure. Because neither of the training groups has acquired new segmental representations, and since none of the stimuli in the repetition priming task have been encountered in the training task, there should be no differences between the two training groups in their priming behaviour. Moreover, *direct-access* models predict the amount of priming to depend on the overall similarity between stimuli. The amount of priming expected in the RELATED condition should therefore be somewhere between that for the IDENTICAL and UNRELATED conditions, since the fricative [ɸ] is more like the fricative [f] than the stop consonants [p, t, k] are.

The predictions of *direct-* and *mediated-access* models differ in two important respects, as highlighted by the ellipses in Figure 8.1. *Mediated-access* models but not *direct-access* models predict there to be a between-group differences in the RELATED condition, and consequently also a *training group* × *priming relationship* interaction. And within each training group, *mediated-access* models predict that two conditions should produce an identical amount of priming: the UNRELATED and RELATED conditions for the PHONEMIC training group, and the RELATED and IDENTICAL condition for the ALLOPHONIC training groups. Note that the predictions for the UNRELATED and IDENTICAL conditions, the two control conditions, are the same.

The within-group predictions of identical amounts of priming are problematic. First, predicting identical performances is always a little problematic as no two groups will behave exactly identically. More importantly, the predictions of *mediated-access* models in Figure 8.1 are based on the assumption that category membership, i.e. whether two speech sound belong

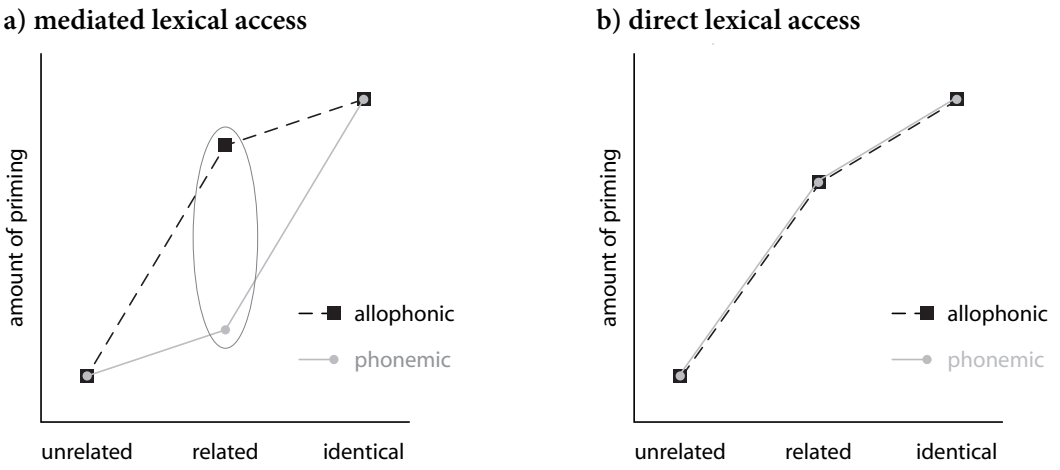


FIGURE 8.2: *Revised* predictions: *mediated-access* (left) and *direct-access* models (right). See running text for further explanation.

to the same or different segmental representations, fully determines repetition priming. This may not be the case; even on a *mediated-access* account, the amount of priming may depend not only on the category membership but also on the *similarity* of segments. For the ALLOPHONIC group this would mean that even though [f] and [ɸ] are treated as members of the same category /f/, because they are physically different, the RELATED condition will produce less priming than the IDENTICAL condition, where primes and probes are not only phonologically but also physically identical. For the PHONEMIC group, because fricatives are more like each other than fricatives and stops are, there may be more priming in the RELATED condition than in the UNRELATED one. This outcome, where priming depends on both category membership and similarity, is illustrated in FIGURE 8.2. Notice that the priming conditions are no longer identical within the training groups; but there is still a between-group difference in the RELATED priming condition, indicated again by the ellipsis. The predictions the *direct-access* models make remain unchanged.

We can conclude that *mediated-access* models make two predictions we have to test. The first is that there will be a *training group* × *priming condition* interaction. If there is, we then need to check whether the ALLOPHONIC group produces more priming in the RELATED condition than the PHONEMIC group. If there is an interaction, but it is the PHONEMIC group that produces more priming, or if the difference occurs not in the RELATED condition but in one of the other two, then *mediated-access* models are not supported.

8.2 The phonetic categorisation task

The PHONEMIC training group has been trained to distinguish /tɪnˈdef/ from /tɪnˈdeɸ/ and /pəˈkif/ from /pəˈkiɸ/. When performing a categorisation task on a [pəˈkif–pəˈkiɸ] continuum (i.e. the OLD continuum), they should respond in a more categorical manner than the ALLOPHONIC group – both *direct-* and *mediated-access* models agree on this. For the NEW continua the predictions of the models disagree; but before we consider these predictions, I need to say what I mean by a performance being *more categorical*.

8.2.1 Defining categoriality

Idealised *categorical* and *continuous* performances are shown in FIGURE 8.3.

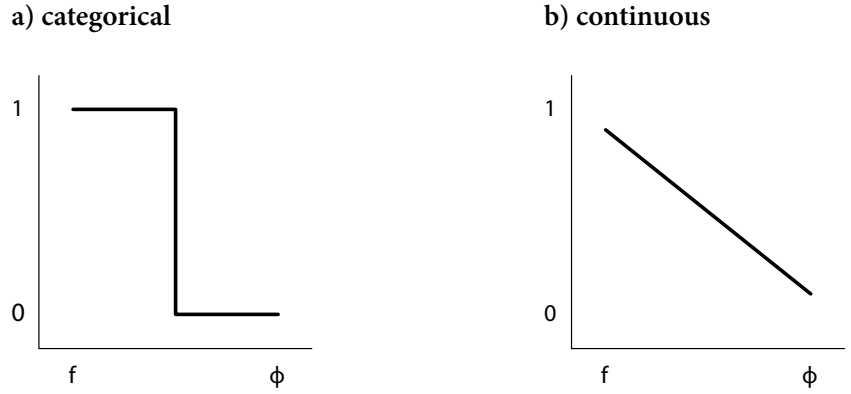


FIGURE 8.3: Ideal categorical and continuous responses in a categorical AXB task. The x-axis represents the continuum (f–ɸ in this case) and the y-axis the proportion of 'F' responses. See text for discussion.

The left-hand graph shows an ideal categorical response curve. All stimuli along the [f–ɸ] are categorised as either being an /f/ or a /ɸ/, and there are no ambiguities. Actual categorisation experiments tend not to produce such a clear-cut categorisation function (see the earlier example in FIGURE 6.1). There are two reasons for this. The first is simply that performance in an experimental task is seldom perfect: subjects make errors. But errors alone cannot explain why in FIGURE 6.1 a), the category boundary is not more like the ideal case in FIGURE 8.3, however; if you go back to FIGURE 6.1 on page 109, you will see that there are two sounds about which subjects are entirely unsure, and two more for which there is some uncertainty. This suggest that, even in a highly categorical case such as VOT, categories are not completely discontinuous, and that we cannot expect categorisation performance to be as perfect as in FIGURE 8.3 a).

The right-hand graph shows an ideal continuous response for the categorical AXB task. The main feature is that, because subjects have no categories for the two speech sounds on which the continuum is based (f–ɸ, in the example), their response ought to be linear and show no

category boundary. Whether this is indeed the case is difficult to say, as categorisation tasks have in general only been used with cross-category and not within-category continua. It depends on the assumption that phonetic similarity is perceived in a linear fashion; and this may not be the case, as the outcome of my phonetic categorisation task suggests, where even the ALLOPHONIC group produced an S-shaped categorisation function (see FIGURE 11.2 on page 166). This outcome could be due to the categorical AXB task, which may encourage a nonlinear categorisation response. In this task, subjects are always presented with the two reference points A and B against which they have to judge the middle sound X: sounds closer to these end points should therefore be categorised more consistently. We can expect this to result in a quite steep categorisation function, as illustrated in FIGURE 8.3 b); but it may also encourage a nonlinear behaviour.[1]

The consequences of all this is that the illustrations of categorical and continuous performances in FIGURE 8.3 should not be taken too literally. The main point is that a categorisation function is *more categorical* if it is more clearly discontinuous or S-shaped than the categorisation function we are comparing it with.

Finally, I should briefly mention how categoriality was measured, or rather how the two training groups were compared. Remember that I analysed the phonetic categorisation data by fitting a mixed-effects logistic regression model (§7.6). If logistic regression curves are fitted individually for each subject, we can take the intercept and slope of these curves as our response features and compare the two training groups with regard to their average intercept and slope.

The meaning of the slope feature should be obvious: a steeper slope indicates a more categorical performance (see again FIGURE 8.3). Whether we should also expect a difference in the intercept is less clear. Even subjects with no category boundaries – i.e. my ALLOPHONIC group – may reach a proportion of 1 and 0 for the end points of the continuum, because if the X is one of the end points, it will be identical to either A or B. But if there is a difference with regard to the intercept, we expect a more categorical performance to have a higher intercept than a more continuous one. This has also been illustrated in FIGURE 8.3, where the continuous categorisation function does not quite reach 0 and 1 at either end (see also FIGURE 8.4 and FIGURE 8.5).

8.2.2 Predictions

As already mentioned, both *direct-* and *mediated-access* models make the same prediction about the OLD continuum. If subjects have successfully completed the training sessions and

[1] Many acoustic parameters are perceived on a nonlinear scale. Intensity and the fundamental frequency f_0, for instance, are both perceived on a logarithmic scale.

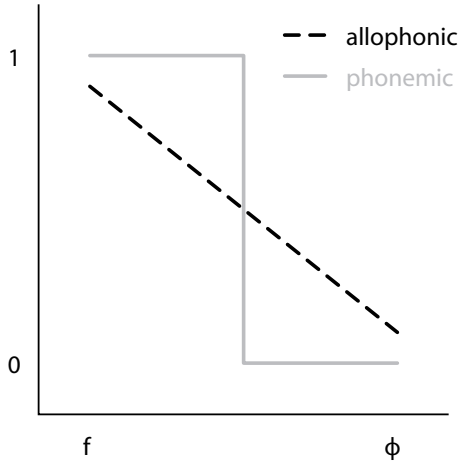


FIGURE 8.4: Idealised predictions for the OLD continuum. Both *direct-* and *mediated-access* models predict that the PHONEMIC group will perform the task more categorically than the ALLOPHONIC group.

can reliably identify the four training stimuli, the PHONEMIC group should perform the categorical AXB task in a more categorical manner than the ALLOPHONIC group, because they are being tested on one of the minimal pairs that they have been trained on.

However, the two types of models differ with regard to the NEW continuum – both the *position* and the *vowel* NEW continuum. The test continua [ˈfelət–ˈɸelət] and [sɑf–sɑɸ] are very different from /tɪnˈdef/, /tɪnˈdeɸ/, /pəˈkif/ and /pəˈkiɸ/, the four training stimuli of the PHONEMIC group. And since on a *direct-access* account all that is acquired are whole words, the four

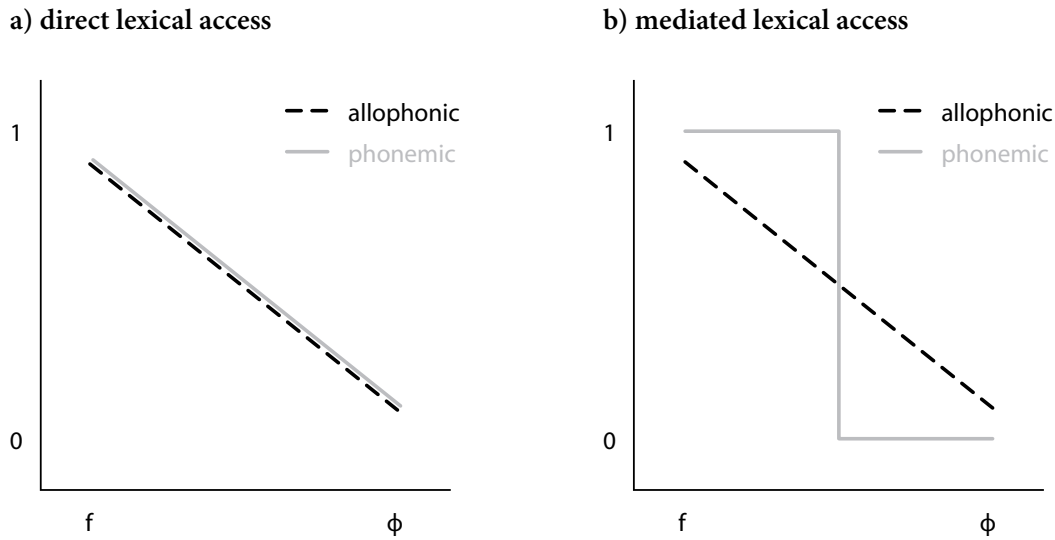


FIGURE 8.5: Idealised predictions of *direct-* and *mediated-access* models regarding the categorical AXB task with the NEW continua.

words learnt by the PHONEMIC group should not have any effect on the processing of the NEW test continua. The consequence of this is that *direct-access* models predict that the PHONEMIC group will be no more categorical than the ALLOPHONIC group with the two NEW continua.

On the *mediated-access* account, if subjects in the PHONEMIC training group have acquired a new segmental representation for the non-native phoneme /ɸ/ – as they need to have if they can successfully distinguish the four training stimuli – they should in principle be able to use this new segmental representation in a novel phonetic context. In other words, subjects in the PHONEMIC training group should be able to apply their segmental representation for /ɸ/ to the test continua [ˈfelət–ˈɸelət] and [sɑf–sɑɸ]. They should thus process these continua in a more categorical manner than the ALLOPHONIC training group.[2]

[2]Notice that as with the repetition priming task, the predictions of *mediated-access* models are not absolute predictions; what is predicted is a *difference* between the two training groups.

9/ Training results

This chapter presents the training data, first for all subjects (§9.1), and then separately for the subjects that received the *position* and *vowel* NEW continua in the phonetic categorisation task (§9.2). The main conclusion will be that the training task has been successful in making subjects in both training groups acquire their novel words.

9.1 The full data set

The purpose of this analysis of the training data is to see whether subjects successfully acquired the words they had been taught. A second point is to assess if, at the end of the training, both groups were equally proficient.

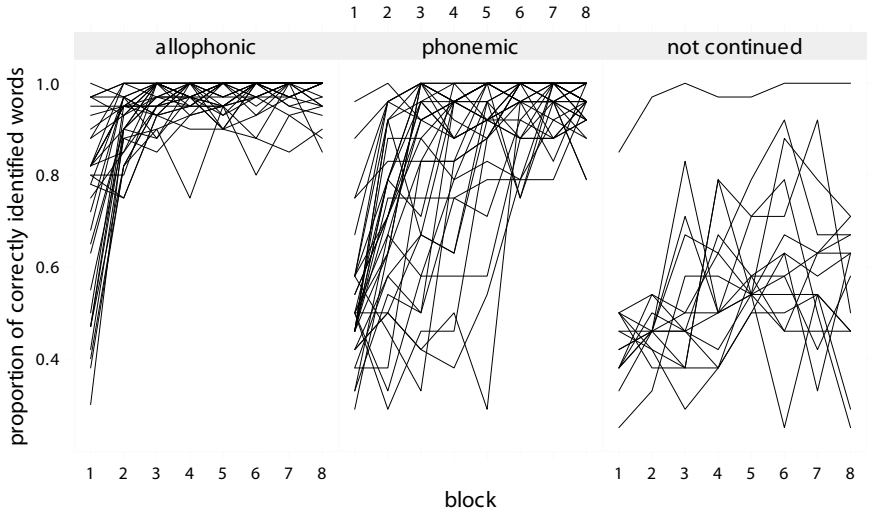


FIGURE 9.1: Performance in the first training session for the two training groups and the participants who did not continue to the second session. The x-axis shows the eight training blocks, and the y-axis the proportion of correct identifications. (The numbers of trials were different for the two groups.)

FIGURE 9.1 shows the performance of the individual participants in the first training session, for both training groups, and also for the participants who did not continue to the second training session. Participants in the ALLOPHONIC group found the task quite easy and generally reached a success rate of 80% or over within a few training blocks. Participants in the PHONEMIC group were learning more slowly overall, but they reached a similar success rate by about the 6th block.

The third panel (on the far right) shows the participants who did not continue to the second training session. The obvious outlier with a very high performance is the one subject from the ALLOPHONIC group who did not return to the second training session. All the others are participants who did not meet the criterion of inclusion; they are all from the PHONEMIC group. The criterion of inclusion was a success rate of at least 80% in one of the training blocks for all four stimuli simultaneously. FIGURE 9.1 shows the overall success rate, i.e. the mean of the four words. But even in this form it can be seen that, with the occasional exception, the performance of participants in the third panel stayed well below the required 80% mark. Overall, there is a clear upward trend even for these subjects.

FIGURE 9.2 shows the performance for the two groups over both training sessions, averaged over subjects.

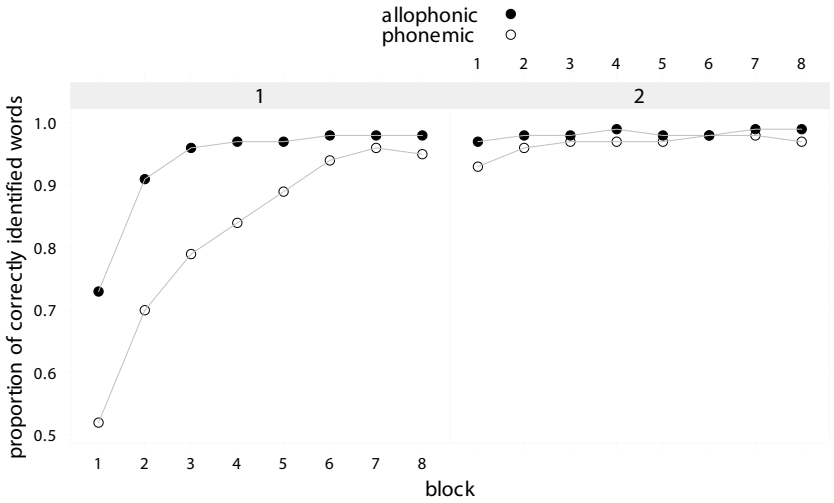


FIGURE 9.2: Performance of the ALLOPHONIC and PHONEMIC training groups in session one (left) and session two (right) of the training. The x-axis shows the eight training blocks, and the y-axis the proportion of correct identifications. Only subjects who took part in both training sessions are included.

Again the much quicker improvement of the ALLOPHONIC group in the first session is obvious; but it is equally apparent that the gap closes in the last three blocks of the first session. Participants' performance at the beginning of the second session was only slightly below that at the

end of the first session, which is evidence for the retention of the newly acquired words between sessions. In fact, no subject performed worse in the second session than in the first, not even those whose performance in the first session was already very high. It also seems that by the second half of the second training session the difference between the groups has virtually disappeared.

To verify this impression, a two-sample test comparing the performance of the training groups in the last four training blocks was carried out. The statistical model was a logistic regression model, i.e. a generalised linear model with a logistic link function and a binomial error distribution. The response variable was the proportion of correct responses, and the factor *group* was the only explanatory variable. Since each subject provides more than one observation, and observations are thus correlated, a random subject term was added to the model. A second reason for using a mixed-effect model is that the design is unbalanced, as the ALLOPHONIC group produced more responses than the PHONEMIC group. The model containing the factor *group* was compared to a null model (without the factor *group*) using a likelihood ratio test. The models were fitted to the data using `lme4`'s Laplace approximation algorithm (Bates and Sarkar, 2007). References for mixed-effects models are given in §7.6.

The test statistics of the likelihood ratio test was $X_1^2 = 1.25$, $p = 0.26$. We may therefore regard the performance of the two training groups in the last four blocks of the second training session as equivalent. To get an idea of the size of this (non-significant) difference, the percentages of correct responses were 98.4% for the ALLOPHONIC and 97.7% for the PHONEMIC group. In absolute terms, a difference of 0.7% would correspond to 0.67 trials for the PHONEMIC group (for whom four practice blocks contained 96 trials) and 1.12 trials for the ALLOPHONIC group (for whom four practice blocks contained 160 trials). This means that on average a subject in the PHONEMIC group made about one error more in the last four blocks than a subject in the ALLOPHONIC group.

9.2 Separate analyses for the two acoustic continua

Two different sets of continua were used in the phonetic categorisation task, with half the participants receiving the POSITION and half the VOWEL continuum (see §5.3 and §7.3). To find out whether, among subjects who got the same categorisation continua, the performance of the two training groups was also equivalent, the test was performed again on the two halves. FIGURE 9.3 shows the performance of the two groups, separately for each of the categorisation continua.

The same *likelihood ratio* test as with the whole data set was carried out on the split data. For

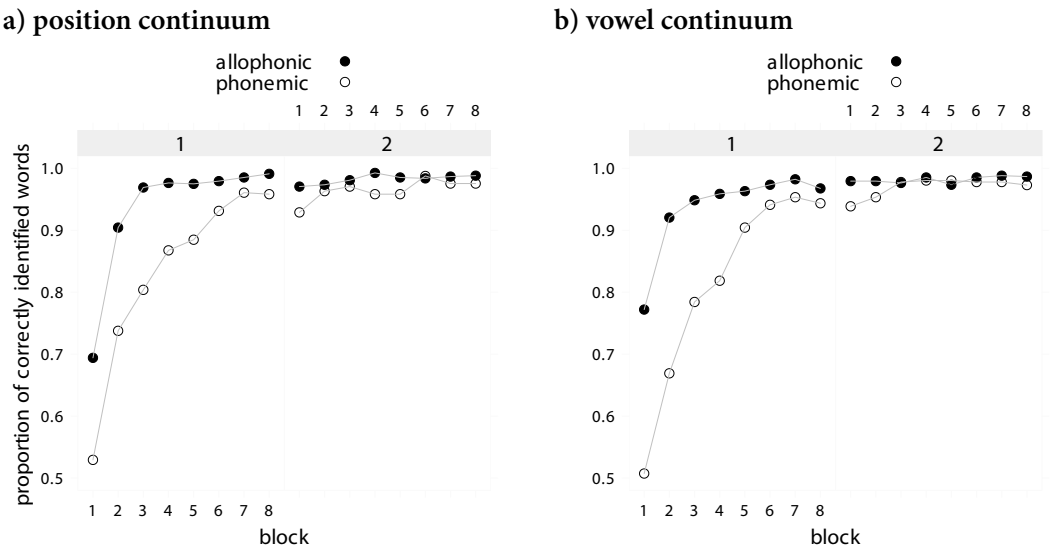


FIGURE 9.3: Performance of the ALLOPHONIC and PHONEMIC training group for the POSITION continua (a) and the VOWEL continua (b). Again, the x-axis shows the eight training blocks, the y-axis the proportion of correct identifications, and the left panel the first and the right panel the second training session.

the POSITION continuum the test statistics is $X_1^2 = 3.05$, $p = 0.08$. And the corresponding mean values are 98.5% for the ALLOPHONIC and 97.5% for the PHONEMIC training group, or a difference of 1%. For the VOWEL continuum, the test statistics is $X_1^2 = 0.01$, $p = 0.9$. And the mean values are 98.3% and 97.8%, respectively, or a difference of 0.5%.

The difference between groups thus seems to be slightly larger for the POSITION continuum than it is for the VOWEL continuum; the difference in block 5 is a likely cause of this difference. But in both cases the null hypothesis of no difference between training groups cannot be rejected. The split data sets thus mirror the full data in that the performance of the two training groups in the last four blocks of the training can be regarded as equivalent.

10/ Repetition priming results

In this chapter on the outcome of the repetition priming test, I will first look at the raw reaction time data in order to screen it for incorrect responses and outliers (§10.1). Then I will analyse the priming data (§10.2), which has been generated by subtracting reaction times to the prime stimulus from reaction times to the corresponding probe stimulus. I will first perform a simple ANOVA and then, after further inspection of the data, a more sophisticated mixed-effects analysis. The outcome will be that the repetition priming data is more consistent with the predictions of *direct-access* models than with those of *mediated-access* models. I will then perform some additional analyses to address certain problems arising from the main analysis (§10.3). My main conclusion will be that it is likely that the repetition priming task has not worked in the way expected, and that therefore its result have to be interpreted with some caution.

10.1 Reaction time data

During the repetition priming task several variables were recorded. The most important is the main response variable, subjects' reaction time (RT) in each lexical decision trial. In addition, I also recorded the correctness of the lexical decision and phoneme monitoring responses, as well as the RTs of the monitoring responses. Information that served to identify the explanatory variables and potential covariates were also logged for each trial: stimulus type (test stimuli, filler stimuli, training stimuli, etc.), the priming relationship of the test stimuli (UNRELATED, RELATED, IDENTICAL), the identity of the stimulus (as listed in Appendix A), stimulus duration in milliseconds, the distance between prime and probe (in terms of the number of intervening stimuli), and finally the training group (ALLOPHONIC, PHONEMIC) that the subject belonged to.

Reaction time was measured from three different points: stimulus onset, stimulus offset, and what I call the *alignment point*. This point was intended to identify where the acoustic information that made it possible to distinguish [f] from [ɸ] and from the stop consonants [p, t, k] became available, and was set at the end of the preceding vowel. It is well known that the

vowel-consonant transition already contains information about the identify of the consonant, but the acoustic signal generally lacks a clear marker of the start of this transition. The end of the vowel, on the other hand, is generally well defined by the disappearance of higher formants in the spectrogram. Note that all of the analyses reported in this section and in §10.2 are based on the RT measurements from the alignment point. The choice of measurement point was not vital to the analysis, however.

Before the reaction time data was transformed into priming data – by subtracting the RT to the probe from the RT to the prime – the raw reaction time data was inspected and screened for outliers.

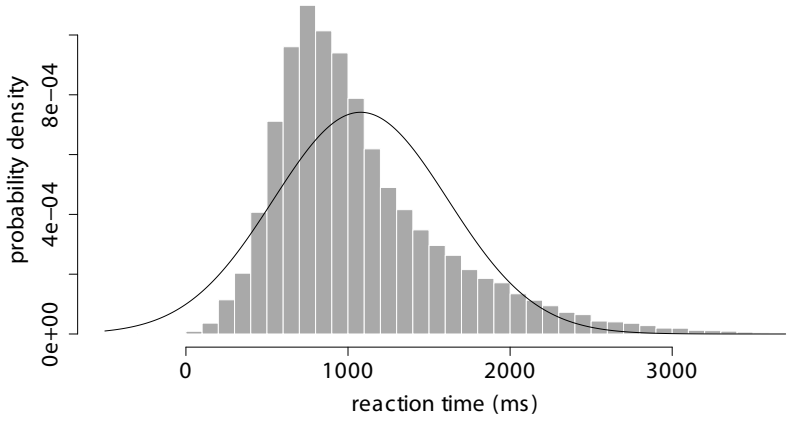


FIGURE 10.1: Histogram of the reaction time data with a normal curve overlaid. RTs from the lexical decision task (measured from the alignment point) are on the x-axis; and the y-axis represents the probability density (i.e. all values sum to 1). The overlaid normal curve has the same mean (1078 ms) and standard deviation (538 ms) as the data.

FIGURE 10.1 shows the overall distribution of reaction times (measured from the alignment point). As we would expect for reaction time measures, the distribution is non-normal. Specifically it is positively skewed, in that the right tail of the distribution (high RT values) is considerably heavier than the left tail (small RT values). In addition, the distribution is also somewhat leptokurtic, i.e. more peaked than we would expect for normally distributed data. The reason for both the skewness and kurtosis is that reaction times are bounded at the lower end: there is a minimum reaction time, but no absolute maximum.

FIGURE 10.2 shows the same data in the form of box-and-whiskers plots (or boxplots for short) for individual subjects. Boxplots provide a rich representation of the location and spread of a variable. The black horizontal lines give the median, the boxes the interquartile range (i.e.

the middle 50% of the distribution), the whiskers extend to 1.5 times the interquartile range from the end of the box, and observations beyond the whiskers are identified as outliers (here shown as circles). FIGURE 10.2 indicates that there is considerable between-subjects variation. Median RT values vary between 662 ms and 1625 ms. There as is also considerable variation in the spread of the data: the interquartile range varies between 328 ms and 946 ms, and the total range between 1323 ms and 3989 ms.

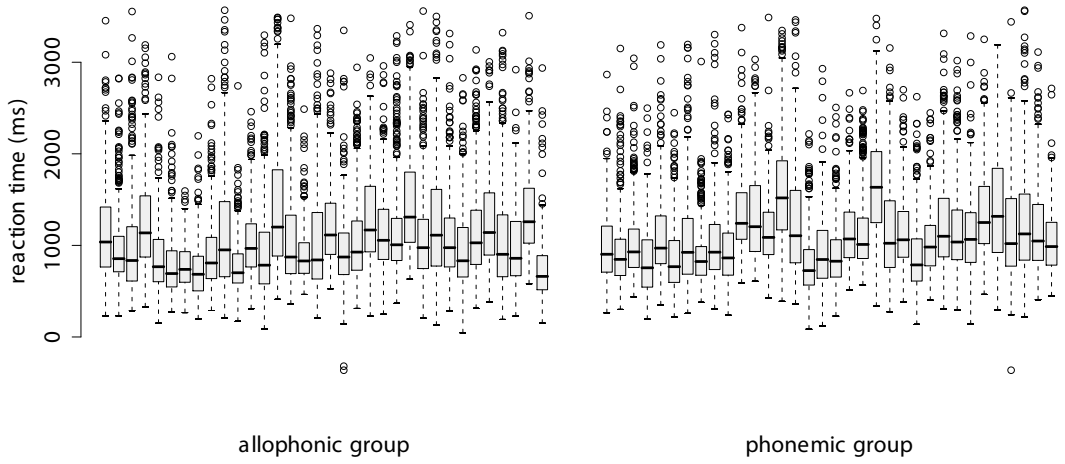


FIGURE 10.2: Boxplots of the reaction time data for all subjects in both training groups. The box represents the interquartile range, the whiskers extend 1.5 times the interquartile range from the end of the boxes, and outliers are identified by small circles. Notice that there is considerable variation between subjects.

With tasks that generate reaction time data, such as lexical decision, it is common to exclude both incorrect responses and outliers. For the purpose of further analysis I prepared three data sets. The first was a *full* data set, where all observations were retained. The second was a *corrected* data set, where only correct lexical decision responses were retained. The third was a *screened* (and corrected) data set, where both incorrect lexical decision responses and outliers were excluded. Outliers were identified and excluded in a two step procedure.

In the first step, all responses with reaction times lower than 200 milliseconds were removed (measured from the alignment point). This follows from what I have said above about reaction times having a lower bound. The cut-off point of 200 ms was chosen because reaction times in decision tasks where subjects have to decide between two responses are typically longer than 200 ms (see Luce, 1959, pp.). It thus seems sensible to assume that a response that is given earlier than 200 ms after the alignment point is likely to result from a decision process that was started earlier. These responses were therefore excluded.

The second step involved the identification of outliers on the basis of the boxplots: only data

points within the whiskers of the boxplots were retained. The interquartile range from which the whiskers are derived is a measure of spread which seems more appropriate for the detection of outliers than the standard deviation, since it does not force normality onto the data, unlike the standard deviation. Outliers were identified individually for each subject, because subjects differed widely both with regard to the location and the spread of their reaction times, as can be seen in FIGURE 10.2. Because of this variation it arguably made more sense to identify outliers for each subject separately, since a common criterion would have excluded almost all responses from one subject and none from another. There still is a place for also using a universal cut-off point, if there is a good reason for it – as there is for the lower bound of 200 ms.

From the three RT data sets (*full*, *corrected* and *screened*) corresponding priming data sets were created by subtracting RT values for the probe from values for the corresponding prime. A Perl script was created for this purpose (reproduced in B.2). The three priming data sets thus created form the basis of all subsequent analyses. The analysis of the priming data in §10.2 focuses on the *screened* set, i.e. the one where both incorrect responses and outliers have been removed.

Of a total of 4080 possible priming responses (68 subjects, and 60 priming pairs each), the *full* data set contains all 4008 responses actually made. In the remaining 72 trials, or 1.8% of all trials, subjects did not give a response in the 4 seconds available. For the *corrected* data set an additional 414 responses, or 10.1%, were excluded because subjects made an incorrect lexical decision to either the prime or the probe stimuli, so that this set contains 3594 responses. The *screened* data set contains 3367 responses, reflecting the removal of an additional 227 responses as outliers.

The distribution of non-responses and excluded responses across the explanatory variables *training group* and PRIMING RELATIONSHIP are as follows for the three sets (the marginals give the row and column totals):

a) *full* data set

	UNRELATED	RELATED	IDENTICAL	
ALLOPHONIC	9	5	10	24
PHONEMIC	20	9	19	48
	29	14	29	72

b) *corrected* data set

	UNRELATED	RELATED	IDENTICAL	
ALLOPHONIC	92	68	72	232
PHONEMIC	110	64	80	254
	202	132	152	486

c) *screened* data set

	UNRELATED	RELATED	IDENTICAL	
ALLOPHONIC	131	121	114	366
PHONEMIC	139	94	114	347
	270	215	228	713

Fisher's Exact Test indicates that the cells in all three contingency tables do not differ significantly from each other (*full:* $p = 1$, *corrected:* $p = 0.56$ and *screened:* $p = 0.21$).

10.2 Priming data: main analysis

As I have explained in §8.1, *direct-* and *mediated-access* models make different predictions about the repetition priming task: only *mediated-access* models predict that there will be a *training group*×*priming relationship* interaction; and that this interaction should be due to the ALLOPHONIC group producing more priming in the RELATED condition than the PHONEMIC group. To test this prediction we need to fit a model that contains the factors *training group* (2 levels: ALLOPHONIC and PHONEMIC) and *priming relationship* (3 levels: UNRELATED, RELATED and IDENTICAL) and their interaction.

In the following, I will first perform a standard by-subjects and by-items ANOVA (§10.2.1). Then I will carry out a more sophisticated analysis by first checking whether the data conforms to the assumptions of linear models, and deciding which covariates to include in the model (§10.2.2); and then by fitting an appropriate mixed-effects model and perform the necessary hypothesis tests (§10.2.3).

10.2.1 Prelude: analysis of variance

The ANOVA, like all subsequent analyses, was performed on the *screened* reaction time data, i.e. the data set where erroneous lexical decision responses and outliers had been removed. Before reporting the ANOVA it is worth comparing means and medians. FIGURE 10.3 shows both a set of parallel boxplots (left) an interaction plot of the priming data by *training group* and *priming relationship* (right).

The boxplots show that there is a slight decrease in the spread of the priming response as we move from the UNRELATED to the RELATED and the IDENTICAL priming condition, particularly for the ALLOPHONIC training group. The notches on the boxes suggest that for the ALLOPHONIC group, but probably not for the PHONEMIC group, the IDENTICAL and RELATED condition are significantly different from the UNRELATED condition.[1]

[1] The width of the notches are approximate 95% confidence intervals for the *median*. The notches are computed as $\pm 1.58 IQR/\sqrt{n}$, where *IQR* is the interquartile range, and *n* the sample size (R Development Core Team, 2006, see entry on `boxplot.stats`; see also McGill et al., 1978). If the notches of two plots do not overlap,

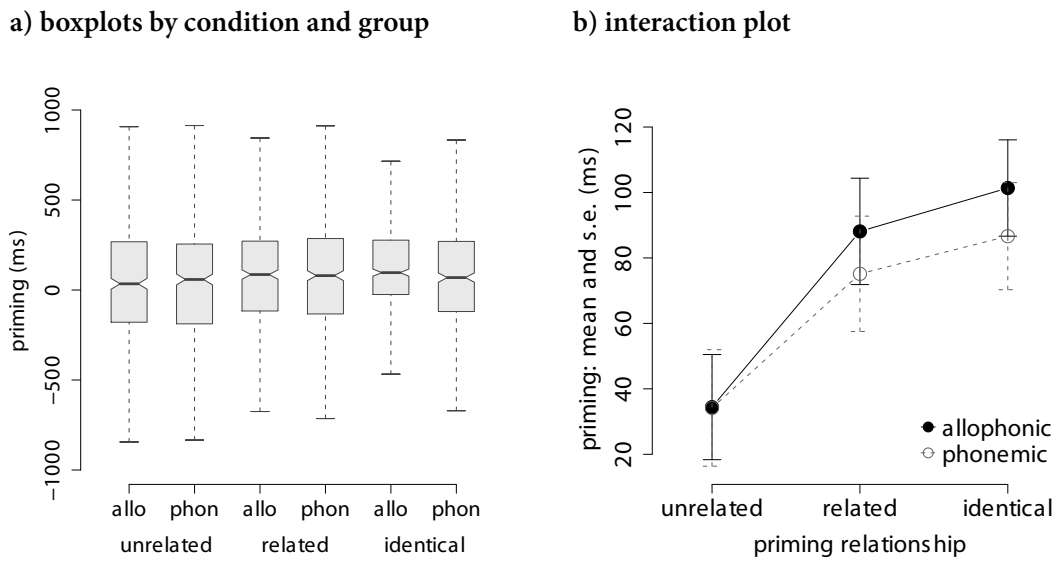


FIGURE 10.3: Boxplots and interaction plot of the screened priming data; both plots compare the two *training groups* at the three levels of the *priming relationship*. The boxplots show the usual median, interquartile range, and whiskers (1.5 times the interquartile range from the edge of the box); the outliers are not shown; the notches correspond to approximate 95% confidence intervals for the medians. The interaction plot gives means and standard errors.

Judging from the interaction plot, it appears that the interaction between the two explanatory variables predicted by the *mediated-access* model did not occur. This impression was confirmed by an ANOVA with *group* as a between-subjects and within-items variable, and *priming* as a within-subjects and within-items variable. *Priming relationship* is the only variable that is significant in both the by-subjects ($F_1 = 6.41$, $p = 0.002$) and by-items analysis ($F_2 = 7.52$, $p < 0.001$). *Training group* is not significant in either analysis ($F_1 = 0.37$, $p = 0.55$; $F_2 = 0.01$, $p = 0.92$), and neither is the interaction ($F_1 = 0.10$, $p = 0.91$; $F_2 = 0.07$, $p = 0.93$).

Pairwise comparisons between the levels of the factor *priming relationship* (RELATED VS. UNRELATED, IDENTICAL VS. UNRELATED and IDENTICAL VS. RELATED) were not carried out, because they do not distinguish between the two models. And pairwise comparisons of the cells of the *group*×*priming* interaction are not warranted, because the interaction was clearly not significant.

the medians of the levels compared can be assumed to be different at the 5% level of significance. The notches thus provide a visual pairwise comparison.

10.2.2 Initial data analysis

We will first check the distributional assumptions of linear models, and then consider covariates for inclusion in the model.

Distributional assumptions

Consider first the histogram on the left-hand side of Figure 10.4. When we compare the empirical distribution with the overlaid normal curve, it is evident that the distribution of the priming data is not completely normal. Unlike for the reaction time data, there is no evidence of a skew, but the distribution is noticeably leptokurtic, i.e. it is more peaked and has heavier tails than a normal distribution.

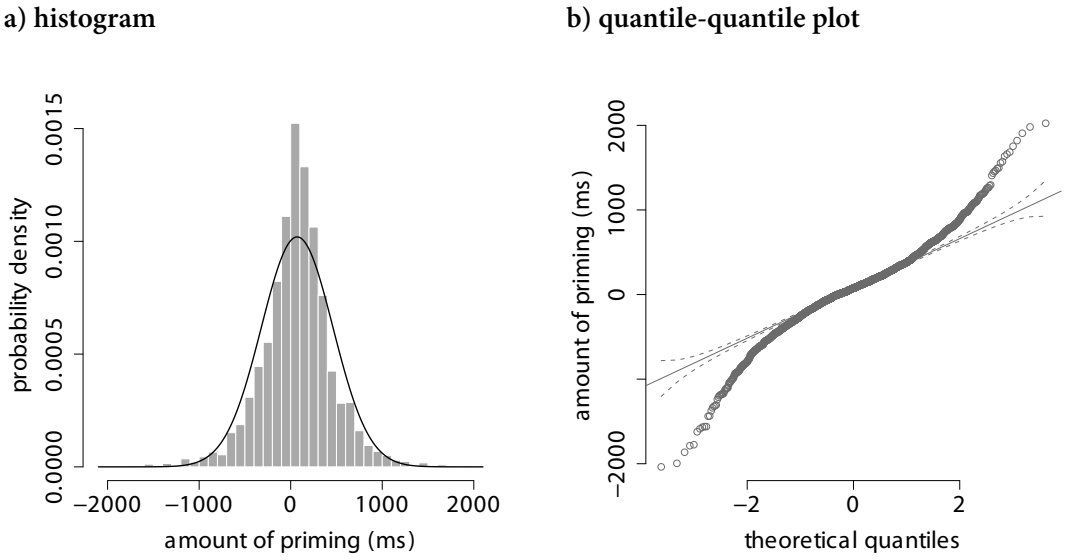


Figure 10.4: Histogram and quantile-quantile (Q-Q) plot of the priming data. In the histogram the x-axis shows the amount of priming (i.e. prime RT minus probe RT), and the y-axis gives the probability density (i.e. all values sum to 1); the overlaid normal curve has the same mean (79 ms) and standard deviation (497 ms) as the data. The Q-Q plot compares the distribution of the priming data (y-axis) with that of a normal distribution (x-axis); the points are the data points plotted against their corresponding theoretical quantiles, the solid line joins the first and third quartiles, and the broken line describes the 95% confidence envelope around the solid line.

The heavy tails show up more clearly in the quantile-quantile (Q-Q) plot on the right-hand side of Figure 10.4. In a Q-Q plot, the quantiles of an empirical distribution are plotted against the quantiles of a reference distribution, a normal distribution in the present case. If the empirical distribution has the same shape as the reference distribution, all points will be on a straight line with a slope of 1. The S-shape we see in Figure 10.4 indicates that the tails of the priming data are farther away from the centre than they would be if the distribution were normal. Should

we be worried by this departure from normality?

Least-squares estimation is generally claimed to be relatively robust against departures from normality – unless some of its other assumptions are also violated, in particular the assumption of constant variance.[2] Constant variance is no problem in the present case (see FIGURE 10.3 again). But it appears that leptokurtosis is the most problematic departure from normality. The reason for this is that with heavy-tailed distributions least-squares estimation is no longer the most efficient method of estimation (Fox, 1997, p. 116, see also Faraway, 2005, 59f.). Alternative methods of estimation include several forms of robust estimation, resampling methods, or the use of theoretical distributions other than the normal.[3] Linear mixed-effects models do not use least-squares but maximum-likelihood estimation; but the above argument should extend to maximum-likelihood estimation, because it makes distributional assumptions which, in the case of linear mixed-effects models, is normality.

None of the alternative methods just mentioned are readily available for mixed-effects models, at the time of writing. I thus decided to instead use the exclusion of outliers as a heuristic strategy. A heavy-tailed distribution can be regarded as distribution with many outliers, and when we exclude these outliers the distribution will become more normal. In the case of repetition priming, we could argue that priming effects as large as 1000 ms or more should be regarded as outliers. They are too extreme to reflect the process of word recognition, given that on average priming rarely exceeds 100 ms. Much larger values are likely to be due to some malfunction or disruption of the process – such as a particularly slow reaction to the prime – even if this kind of malfunction may be relatively common.

A precise cut-off point cannot be established on substantive grounds, and any choice would have to be arbitrary to a certain extent. I therefore used the same method of outlier identification as I used on the raw reaction time data: all observations were discarded which were more than 1.5 times the interquartile range from either end of the interquartile range. In addition, I also kept the original data set for comparison. This is what I meant by saying that the exclusion of outliers was a heuristic strategy: if both sets produce equivalent results we can have more trust in them than if we had only considered the original, leptokurtic data set.

FIGURE 10.5 shows an interaction plot of both the original priming data (on the left, repeated

[2] The main assumptions of least-squares estimation are *linearity* (the expected value of the dependent variable is a linear function of the independent variable), *independence* (observations are sampled independently) and *constant variance* (the error variance does not depend on the values of the independent variables). The important Gauss-Markov theorem, which demonstrates that the least-squares estimator is the most efficient of all linear unbiased estimators, follows from these three assumptions. (An efficient estimator has less variance than a less efficient estimator.) If the data are also normally distributed, least-squares estimators are the most efficient among *all* estimators. See e.g. Fox (1997, pp. 112–118) or Faraway (2005, 13–16).

[3] See e.g. weFaraway (2005, 59f.). With regard to the use of other types of distribution, Rouder et al. (2005) have suggested using Weibull distributions to model reaction time data.

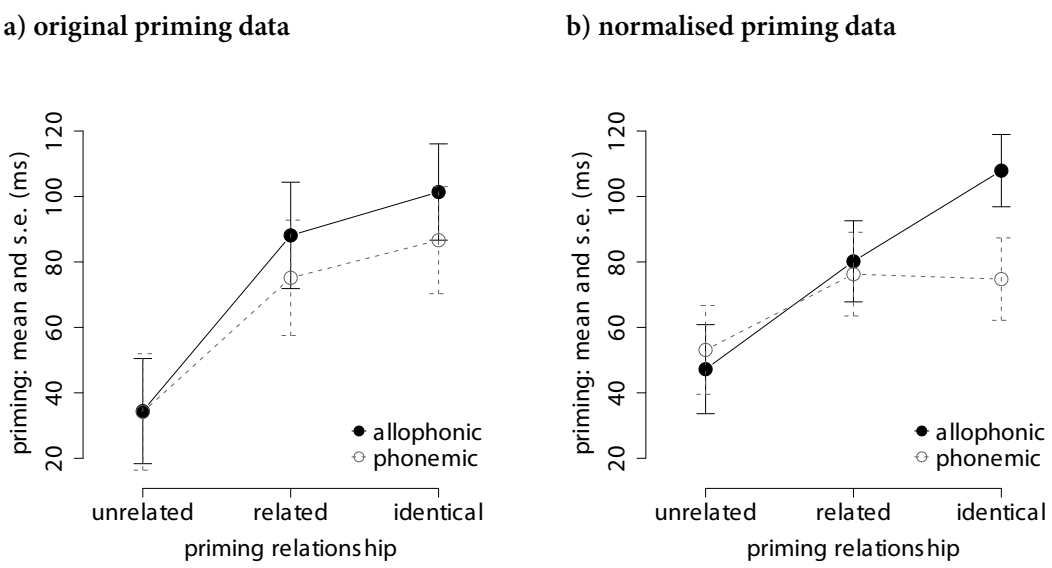


FIGURE 10.5: Interaction plots: mean priming for the two *training groups* at the three levels of *priming relationship*. Left-hand plot: priming data previous to exclusion of outliers; right-hand plot: the same data after exclusion of outliers. The error bars give standard errors.

from FIGURE 10.3) and the new 'normalised' data (on the right). The interaction plot for the normalised data set seems to confirm that the interaction effect predicted by the *mediated-access* model – significantly more facilitation for the ALLOPHONIC group in the RELATED condition – is not supported. It also indicates however, that there *may* be a group difference in the IDENTICAL condition. This outcome was not predicted by either model and would be difficult to interpret, because comparable outcomes in the two control conditions (UNRELATED and IDENTICAL) was an essential assumptions when designing the repetition priming task.

Covariates

In our case, the main purpose of including covariates was to reduce the residual variance in the model and thus to make the subsequent hypothesis tests more powerful. The two types of word recognition models make no predictions about these covariates; no significance test were therefore carried out on the covariates, apart from goodness-of-fit tests to help us decide whether a covariate should be included in the model.

The following covariates were considered for inclusion:[4]

1) *position*: the point within a test list where an observation was made, or more precisely the position where the prime stimulus occurred;

[4]Note that the set of potential covariates was chosen before the experiments were run, so that their values could be recorded during the experiments.

2) *distance*: the distance between prime and probe (8,10,12 or 14 steps);

3) *prime duration*: the duration in milliseconds of the prime stimulus;

4) *probe duration*: the duration in milliseconds of the probe stimulus;

5) *monitoring*: the proportion of correct monitoring responses made by the subject.

A covariate was included in the model if its inclusion improved the fit of the model by $p = 0.1$. This relatively lenient (but common) criterion was chosen because, when in doubt, it is better to include a covariate than to leave it out; even marginally improving the fit of the overall model will result in a more powerful test of the experimental hypotheses. Goodness-of-fit was established by the usual likelihood ratio test statistics.

Position ($X_1^2 = 17.4, p < 0.001$), *probe duration* ($X_1^2 = 18.3, p < 0.001$) and its square ($X_1^2 = 3.25, p < 0.008$) clearly increased the fit. The square of the *position* was a borderline case ($X_1^2 = 2.74, p < 0.098$), but a parametric bootstrap with 1000 simulations suggested that $p = 0.068$ is a more precise p-value, and that the square of *position* should also be included in the model.[5] *Prime duration* was also significance, but it adds nothing to the fit of a model where *probe duration* is already included; and since *probe duration* accounted for more variation than *prime duration*, it was the former that was included as a covariate. None of the other variable resulted in an improvement of the fit.

FIGURE 10.6 presents scatterplots of *priming* against *position* (left) and *probe duration* (right). The broken lines are least-squares regression lines. The negative slope for *position* indicates that the amount of priming diminishes as subjects progress through the task, and the positive slope for *probe duration* that longer stimuli produce more priming. The solid lines are lowess non-parametric regression lines.[6] They indicate that the effects are non-linear, and probably quadratic.

The effect of *position* seems to level off at about position 250: after that point the amount priming no longer diminishes. If we assume that the reduction in priming is an effect of practice – which is supported by the fact that RTS also decrease over the course of the priming experiment – the levelling-off could mean that subjects have reached their peak performance at this point. It could also mean that the effect of practice is counteracted by an effect of fatigue. In any case, the nonlinearity of the positional effect suggests that we should also include the square of *position* in the model.

[5] Note that the likelihood ratio test statistics used to compare mixed-effects models only has an approximate χ^2 distribution, and that a parametric bootstrap (which simulates the responses under the null hypothesis) is recommended in critical cases (Faraway, 2005, pp. 158ff.)

[6] A lowess non-parametric regression line is a local regression line: for each data point, observations closer to it are given more weight than observations that are farther away. The result is a non-parametric regression line. Depending on how the weights are skewed towards the centre, lowess lines can be relatively smooth, as in FIGURE 10.6, or more discontinuous.

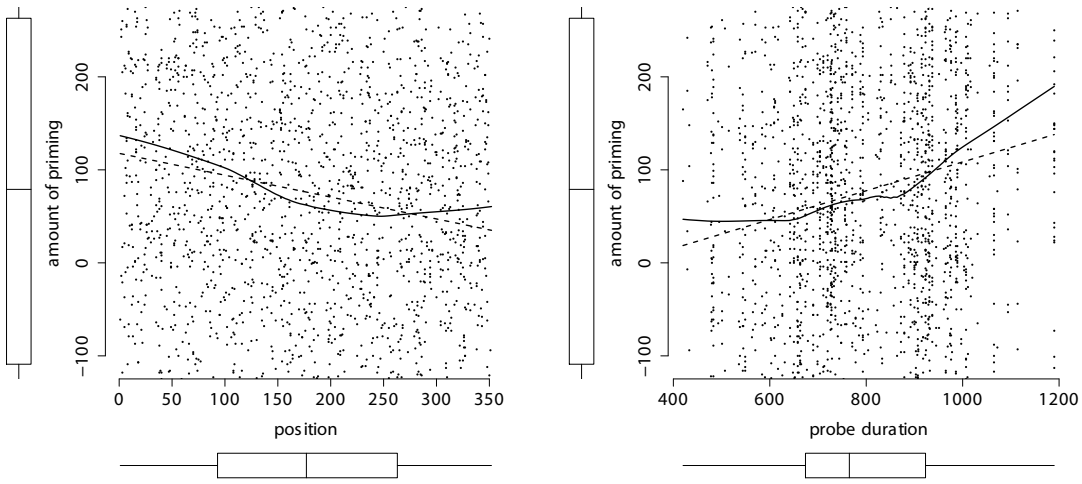


FIGURE 10.6: Scatterplots of the response variable (priming in milliseconds) against *position* (left) and *probe duration* (right); with marginal boxplots, least-squares (broken) and lowess non-parametric regression lines (solid). Notice that the range of the y-axis has been truncated so as to focus on the middle 50% of the priming distribution. See running text for discussion.

The effect of *probe duration* seems to increase more or less steadily from a slope of about zero at low durations until it reaches a uniform slop of about 0.37 after a probe duration of 900 ms; from that point onward each increase of the probe duration of 10 ms leads to an increase in priming of 3.7 ms. It is hardly surprising that priming increases with stimulus duration: the longer the probe, the more time there is to respond early, and the more facilitation we expect. The zero slope with short stimuli suggests that there exists a baseline of priming of about 50 ms. The upwards bend at a probe duration of 900 ms may have to do with the difference between mono- and disyllabic stimuli. It may be that the tripartite shape of the non-parametric regression line is due to this difference: all stimuli shorter than 609 ms are monosyllabic (the mean duration of monosyllables is 654 ms); between 609 and 997 ms we find both mono- and disyllabic stimuli; and above 997 ms all stimuli are disyllabic (the mean duration of disyllables is 897 ms); these points roughly correspond to the bends in the scatterplot. This could be taken to mean that stimulus duration only has an effect on priming when stimuli are disyllabic but not when they are monosyllabic, as the line is horizontal for the purely monosyllabic stimuli. Whatever the reason for the bends in the lowess regression line, it suggests that both *probe duration* and the square of *probe duration* should be included as covariates in the model.

10.2.3 Model diagnostics and hypothesis testing

After the inclusion of covariates, the model contained fixed factors for *training group*, *priming relationship*, their interaction, *position* and the square of *position*, and *probe duration* and the square of *probe duration*; in addition to the random factors for *subject* and *item*.

a) before removal of outliers **b) after removal of outliers**

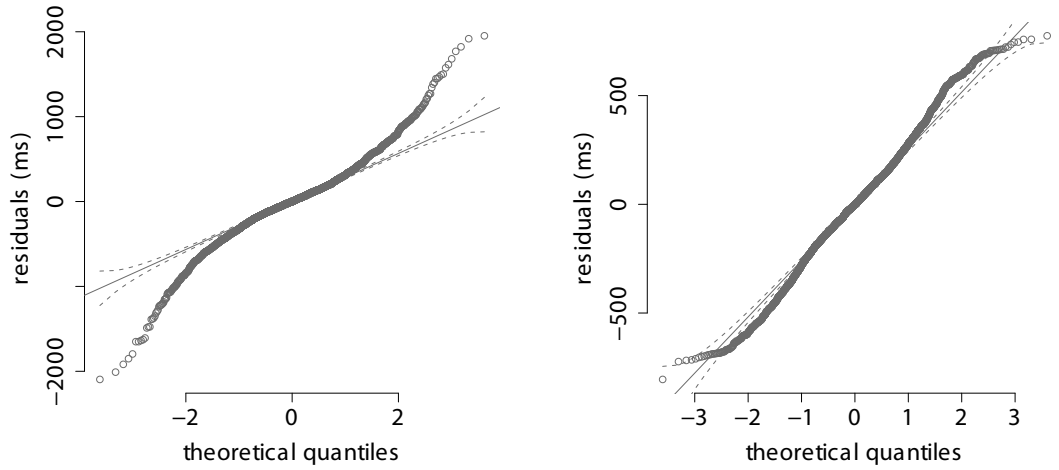


FIGURE 10.7: Quantile-quantile (Q-Q) plots of the priming data before (left) and after (right) the removal of outliers. The solid line joins the first and third quartiles, and the broken line give a 95% confidence envelope around it.

FIGURE 10.7 presents Q-Q plots of the residuals. The one on the left, plots the residuals of the model fitted to the original data against a normal distribution, and the right-hand figure the residuals of the same model fitted to the normalised data. The first Q-Q plot has very similar shape to the Q-Q plot of the response variable, as given in FIGURE 10.4. This is to be expected; the distribution of the residuals of a model would only be very different from the distribution of the response variable, if one of the explanatory variables included in the model accounted for most of the outliers. The second Q-Q plot indicates that the distribution of the residuals of the normalised data set is much closer to normal – which was the purpose of the exclusion of outliers. There is still some departure from normality in the tails of the distribution, but the departure is considerably smaller than it is in the original data set.

Hypothesis tests were carried out on both data sets in the following order: (i) the *training group*×*priming relationship* interaction, (ii) the main effects for *training group*, and (iii) *priming relationship*. The outcomes of the tests were as follows:

1) The *training group*×*priming* interaction was not significant, neither with the normalised

data set ($X_2^2 = 2.89$, $p = 0.24$) nor with the original data set ($X_2^2 = 0.31$, $p = 0.86$).

2) The factor *training group* was also not significant, neither with the normalised ($X_1^2 = 0.85$, $p = 0.36$) nor with the original data set ($X_1^2 = 0.46$, $p = 0.5$).

3) The factor *priming relationship* was significant with both sets: $X_2^2 = 7.35$, $p = 0.025$ with the normalised and $X_2^2 = 10.2$, $p = 0.0063$ with the original set.

While there were small differences between the two data sets, the outcome is very clear (and consistent with the earlier ANOVA): there is no evidence for the important *training group*×*priming* interaction predicted by *mediated-access* models, but a robust effect of *priming relationship*.

Even though the interaction was not significant, because *mediated-access* models predict a group difference in the related condition, a pairwise comparison of the two training groups in the RELATED condition was carried out using the MCMC method mentioned in §7.6. This difference was clearly not significant, neither with the normalised ($p = 0.66$) nor the original data set ($p = 0.65$). In addition, I also carried out a pairwise comparison of the training groups in the IDENTICAL condition (with the normalised data set only, where a difference seems possible). This difference was also not significant ($p = 0.10$); and even if it had been, because this comparison was not planned but was suggested by the data, the outcome would have to be treated with caution.

10.3 Some additional analyses

As we have just concluded, there was no evidence for a *training group*×*priming relationship* interaction that was caused by a between-group difference in the RELATED condition. This outcome of the repetition priming test is consistent with the predictions of *direct-access* models but not with the predictions of *mediated-access* models. But *direct-access* models are only supported by a null result; and it is of course possible that the non-occurrence of an effect has another reason than the one assumed. In this brief section, I want to consider the question whether there is any evidence which should make us cautious to interpret the null result for the *training group*×*priming condition* interaction as support for the *mediated-access* model. This issue will be taken up again in more detail in Chapter 12.

First, the normalised data set (see FIGURE 10.5 again) suggests that there could a difference in the IDENTICAL condition. This difference is not significant at the .05 level; but with $p = 0.10$ it has a quite low probability of having occurred by chance, and with size of 33 ms it would be quite a large difference. If this difference in the IDENTICAL condition were genuine, it would

violate one of the basic assumptions of the experiment, namely that the IDENTICAL condition can serve as a control which defines the ceiling for facilitation. The problem would be that because the two groups have different ceilings, the values for the RELATED condition can no longer be compared across the two training groups.

A second problem regards the definition of facilitation. The standard assumptions is that the repetition of a stimulus facilitates the processing of this stimulus, which makes reaction times to probe stimuli faster than reaction times to prime stimuli. But a similar difference between primes and probes could, in principle, also occur if a participant was for some reason not faster with the probe but slower with the prime. In particular, if one of the training groups were in general slower at processing prime stimuli, this would show up in the data as increased facilitation compared to the other group. We should thus compare groups with regard to their processing of prime and probe stimuli.

To make such a comparison possible, we can make use of the fact that both groups had an almost identical performance in the UNRELATED condition; the UNRELATED condition can thus serve as a baseline for the comparison. If we subtract RTs in the UNRELATED condition from RTs in the other conditions, we get the following means for the prime and probe stimuli (a negative value now indicates faster performance):

Primes	IDENTICAL	RELATED
ALLOPHONIC	−38	−21
PHONEMIC	−32	14

Probes	IDENTICAL	RELATED
ALLOPHONIC	−105	−74
PHONEMIC	−60	−26

Note that primes in the IDENTICAL condition – which were simply repeated as probes – were processed faster than primes in the UNRELATED condition. The reason for this could be that all primes in the IDENTICAL condition ended in the labiodental fricative [f], while primes in the UNRELATED condition ended in either [f] or one of the stop consonants [p, t, k]. Given that stimuli ending in [f] were the most common, such an overall faster performance could just be a result of subjects getting used to them.

In the RELATED condition, however, the ALLOPHONIC group but not the PHONEMIC group was somewhat faster in processing the prime stimuli, again compared to the UNRELATED condition. When looking at the reaction times to the probe stimuli we find a similar picture: subjects in the ALLOPHONIC training group were faster at processing the stimuli than subjects in the PHONEMIC group. The reason for this may be that in the training the PHONEMIC group has learnt to pay more attention to the subtle differences between [f] and [ɸ], and that this increased attention now slows down their processing, particularly in the RELATED condition where stimuli that end in [f] and [ɸ] occur.

10.4 Conclusions

The main conclusion from the repetition priming test is that there is no *training group*×*priming relationship* interaction, and also no difference between training groups in the RELATED priming condition. This outcome contradicts the predictions of *mediated-access* models and is consistent with *direct-access* models.

However, there are some doubts about whether these result can be taken at face value; partly because there was some suggestion of a group difference in the IDENTICAL control condition, and partly because subjects in the PHONEMIC group may have performed the test task in a different manner from subjects in the ALLOPHONIC group. These issues will be discussed further in Chapter 12.

11 / Phonetic categorisation results

In this chapter, I will present the outcome of the phonetic categorisation task. After a qualitative analysis in §11.1, I will fit a logistic regression model to the data (§11.2.1), and then perform a response feature analysis of the slope and intercept of the logistic regression model (§11.2.2). The main conclusion will be that one of the two NEW continua (the *vowel* continuum) provides support for *mediated-access* models. The consequences of this outcome will be further discussed in Chapter 12.

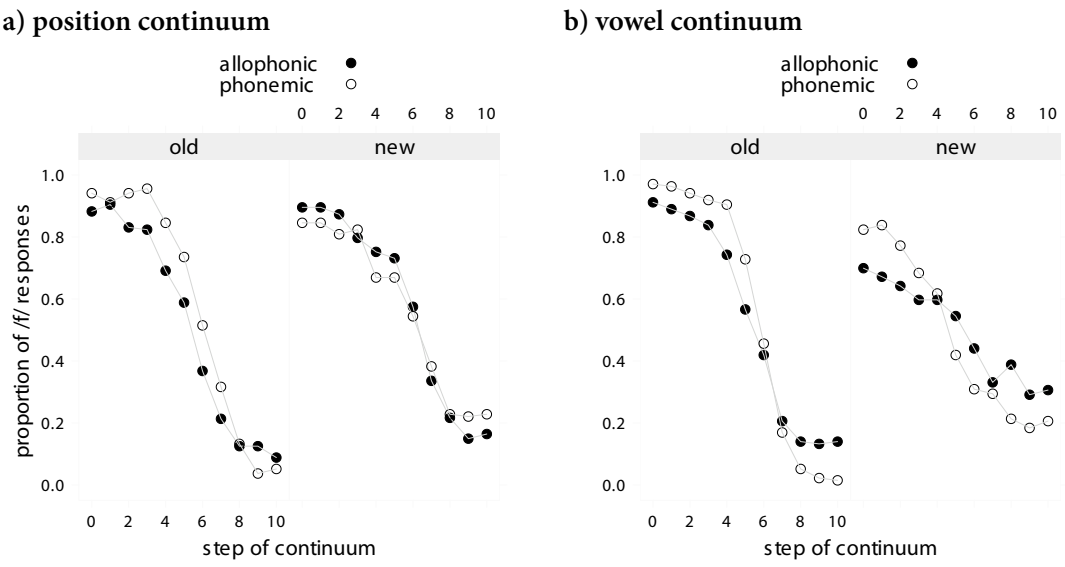


FIGURE 11.1: Categorisation functions for the *position* and *vowel* continua. The x-axis shows the 11 sounds of the continuum and the y-axis the proportion of /f/-responses. The continua were [pəˈkif–pəˈkiɸ] for both OLD continua, and [ˈfelət–ˈɸelət] (left) and [sɑf–sɑɸ] (right) for the NEW continua.

11.1 Qualitative analysis

The categorisation functions for the *position* and *vowel* continuum are shown in FIGURE 11.1. Compare these with the predictions that *direct-* and *mediated-access* models make about the NEW continua in FIGURE 11.2 (repeated from FIGURE 8.5).

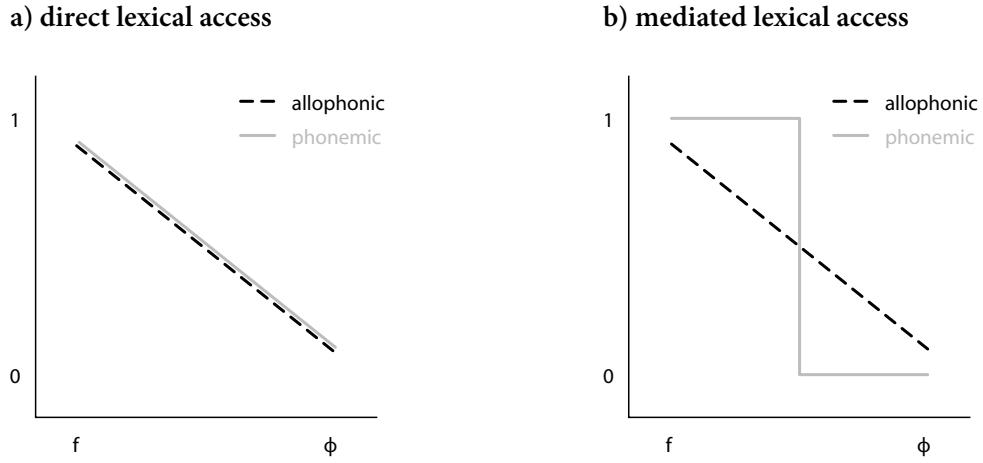


FIGURE 11.2: Idealised predictions of *mediated-* and *direct-access* models regarding the NEW continuum. Compare this to the results in FIGURE 11.1.

The OLD continua present clear evidence that the PHONEMIC group performed in a more categorical manner than the ALLOPHONIC group. The categorisation functions of both groups are fairly categorical, but the function of the PHONEMIC group is more categorical than that of the ALLOPHONIC group: it is more extreme at either end and has a steeper category boundary. For the *vowel* NEW continuum (right-hand graph), we find that both functions are much less categorical than for the other continua; but there is a clear difference between the training groups, with the PHONEMIC function again being more categorical than the ALLOPHONIC function. For the *position* NEW continuum (left-hand graph), there is no difference between the category boundaries or slopes of the two functions. However, we find a slight between-groups difference towards the end points of the continuum; but this difference goes in the opposite direction from that predicted by the *mediated-access* model: it is the ALLOPHONIC function that seems the more categorical.

The visual inspection of the categorisation function thus suggests that the *mediated-access* model is correct for the *vowel* continuum, and the *direct-access* model for the *position* continuum. Before verifying this impression quantitatively, I briefly consider what could have caused the *vowel* continuum to produce a less categorical performance overall than the *position* continuum. One possibility is that it is due to the difference in the number of syllables. The *position* continuum was also disyllabic like all the training stimuli, while the *vowel* continuum was

monosyllabic. We would then have to say that a change in numbers of syllables results in an overall reduction of categoriality, but that keeping the position of the [f–ɸ] contrast constant between training and test results in a categoriality *difference* between the two training groups.

11.2 Quantitative analyses

11.2.1 A logistic model

There are two explanatory variables: the training *group*, and the *steps* of the continuum. First, I want to argue that *novelty* (i.e. OLD vs. NEW) needs not be treated as additional variables, because the predictions that the two models make for the different continua can be regarded as independent. For the OLD continuum, both models make the same prediction: that the PHONEMIC group is more categorical than ALLOPHONIC group. For the NEW continuum, on the other hand, only the MEDIATED-ACCESS model predicts a *group* difference. Nevertheless, because the analysis shows that the two continua behaved differently, I will perform an analysis with *novelty* included at the very end of this chapter (see p. 170).

As a first attempt to test the predictions of the models, we can look for a *group*×*step* interaction. The main effects are irrelevant. A main effect of *step* would mean that there is a difference in the proportion of /f/-responses between the steps of the continuum. We expect this effect to be significant, as it would only be non-significant if the identification curves where approximately horizontal. A main effect of *group* – which we might think to be relevant – would not be a good indicator of a differences in *categoriality*. We would get a main effect of *group* if the category boundaries of the two groups were in a different place, or if the categorisation function of one of the groups were closer to either of the extremes (say closer to 1) throughout most of the continuum. But if the difference is only with regard to categoriality and nothing else, there will be *no* main effect of *group*, because what the more categorical function is closer to 1 at the left end it will be closer to 0 at the right end.

It is thus only a *group*×*step* interaction that, in a first approximation, can be used as a test of a differences in categoriality between the training groups. Note however, that this is a rather poor approximation of the concept of categoriality, because many kinds of differences between the shape of the two curves would result in a significant interaction effect. A response feature analysis is more appropriate, as we shall see in §11.2.2.

The response variable is binary: /f/ or /ɸ/. The appropriate model is a logistic regression model, i.e. a generalised linear model with a logit (or logistic) link function and a binomial distribution.[1] The explanatory variables were *group*, as a categorical variable, and *step*, as con-

[1] As I have mentioned before (§9.1), other link functions such as a probit link would also be appropriate. I have

tinuous variable. The within-subject correlation was taken into account with a random factor for *subject*. As the response variable we could either have used all the binary responses or the proportions that we have already computed for the categorisation functions. I chose the second option, because it was easier to compute. The test performed was the usual likelihood ratio test; and the estimation method used was `lme4`'s Laplace estimation algorithm (Bates and Sarkar, 2007).

All interactions were highly significant, for both OLD continua (*position*: $X_1^2 = 224, p < 0.001$; *vowel*: $X_1^2 = 627, p < 0.001$), and both NEW continua (*position*: $X_1^2 = 77.1, p < 0.001$; *vowel*: $X_1^2 = 328, p < 0.001$). What this outcome mainly suggests it that simply testing for an interaction is not specific enough if we want to assess categoriality. So let us turn to a more meaningful response feature analysis.

11.2.2 Response feature analysis

To perform this analysis, a regression line, a categorisation function in other words, was fitted separately for each subject. The intercepts and slopes of these regression lines could then be used as the new response variables. The intercept is a measure of how closely the categorisation function approaches 1 at the /f/-end of the continuum. It can thus be taken as a measure of categoriality: a higher intercept indicates a more categorical performance. The slope represents the steepness of the function, and is a particularly meaningful indicator of categoriality (see §8.2.1): the more negative the slope, the more categorical the performance.

The model fitted to subjects' responses was a logistic regression model as before, but this time with *step* as the only explanatory variable. The factor *group* and the random effect for subjects are obviously not meaningful, as *group* is a between-subjects variable, and as there is no between-subjects variation in the data of an individual subject.

Let us look at the intercepts first. Boxplots[2] of the intercepts, transformed back from logarithmic odds to proportions, are presented in FIGURE 11.3. The effect of the training is very obvious for the OLD continuum. The PHONEMIC training group, with both the *position* and *vowel* continuum, has a median close to 1 and a very narrow interquartile range; the ALLOPHONIC training group has a lower median intercept and also shows more spread. With the NEW continua there is more spread overall – which is to be expected as there was no training for these continua. But for the *vowel* continuum, the PHONEMIC group has again a higher intercept.

chosen the logistic function for its mathematical simplicity and ease of interpretation.

[2] These plots show the median (heavy line), and give an indication of the spread of the data: box = interquartile range; whiskers = 1.5 times the interquartile range from the edge of the box; circles = outliers).

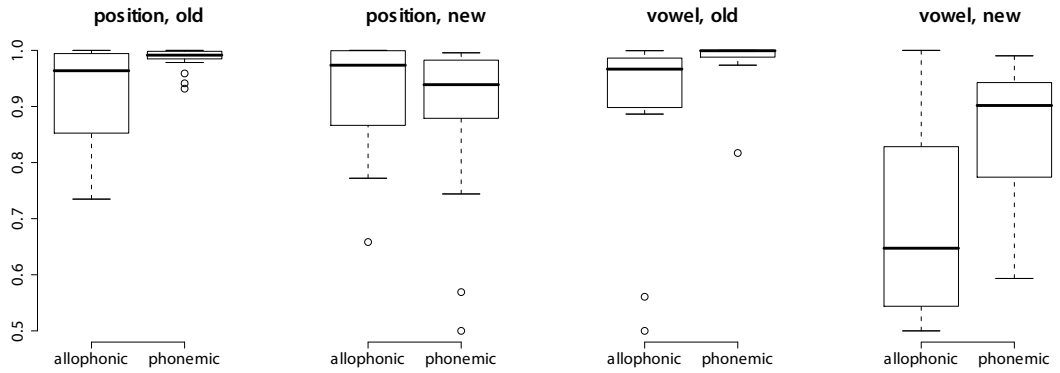


FIGURE 11.3: Boxplots of the intercepts of the fitted categorisation functions. The two plots on the left are for the *position* continuum and the two on the right for the *vowel* continuum. The y-axis represents the intercept of the fitted categorisation functions as a proportion of /f/-responses.

Since proportional data is non-normal – as can be seen from the boxplots – I decided to perform a non-parametric Wilcoxon rank-sum test (also known as a Mann-Whitney test) on the intercept data. The null hypothesis was that there is no difference between training groups. The test was carried out one-sided, because the predetermined alternative hypothesis was that the PHONEMIC group is more categorical (and therefore produces higher intercepts) than the ALLOPHONIC group. The difference for the OLD continuum for both the *position* ($p = 0.016$) and *vowel* ($p < 0.001$) test groups is significant, as is the difference for the *vowel* NEW continuum ($p = 0.008$). For the *position* NEW continuum, the difference is not significant ($p = 0.884$). It is obvious from both the boxplots and from FIGURE 11.1 that with the *position* continuum, it was the ALLOPHONIC training group which has a slightly higher intercept. To test for this possibility I reversed the test; but the reverse difference was also not significant ($p = 0.123$).

The interpretation of the slope is straightforward: the steeper the slope (i.e. the more negative its value) the more categorical the categorisation function. The boxplots in FIGURE 11.4 show that the median slope of the PHONEMIC group is steeper than that of the ALLOPHONIC group, again with the exception of the *position* NEW continuum. Note that in this case the direction of the y-axis is reversed so that steeper slopes are higher up on the axis, and the representation is in log odds.

I again carried out a one-sided Wilcoxon rank-sum test. The OLD continuum is again different fort both test groups (*position*: $p = 0.037$; *vowel*: $p < 0.001$). The same is the case for the *vowel* NEW continuum ($p = 0.009$), but not for the *position* NEW continuum ($p = 0.909$). If we again reverse the direction of test for the *position* continuum, it remains non-significant ($p = 0.097$).

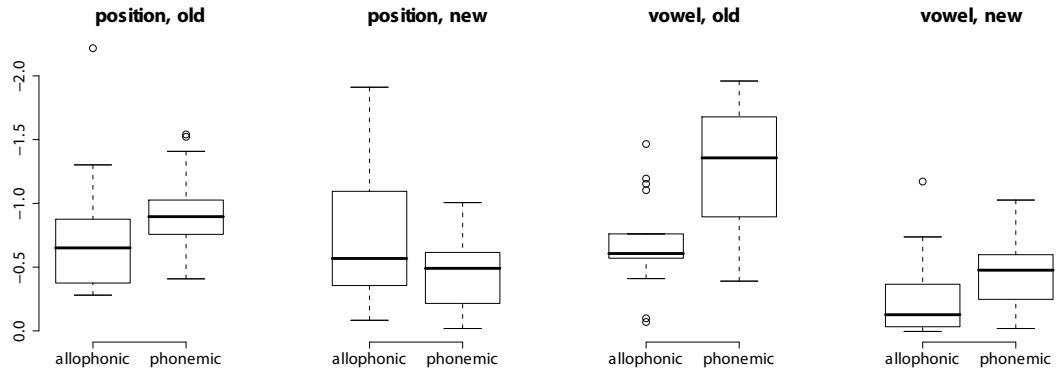


FIGURE 11.4: Boxplots of the slopes of the fitted categorisation functions. The two plots on the left are for the *position* continuum and the two on the right for the *vowel* continuum. The y-axis represents the slope of the fitted categorisation functions in log odds; the more negative the value the steeper the slope.

I have argued earlier that the OLD and NEW continua should be regarded as independent experiments and that *novelty* need therefore not be included as an explanatory variable. This argument still holds in principle; but because the results were different for the two types of continua – a group difference for the *vowel* but not the *position* NEW continuum – it is worthwhile to perform a three-way ANOVA with the factors *novelty* (within subjects) *continuum* (between subjects) and *group* (between subjects) in order to judge whether the difference between the two continua was reliable. Because of the non-normality of the distribution of both the intercepts and slopes, the coefficients of the least-squares ANOVA model were compared with those of a robust model based on M-estimation (Faraway, 2005, pp. 98–101).

For the intercepts, the three-way interaction was not significant ($F(1, 128) = 0.06$, $p = 0.81$) but all of the two-way interactions were (*novelty*×*continuum*: $F(1, 128) = 9.98$, $p = 0.002$; *novelty*×*group*: $F(1, 128) = 13.79$, $p < 0.001$; *continuum*×*group*: $F(1, 128) = 7.38$, $p = 0.008$). The pattern was the same for the slopes (three-way interaction: $F(1, 128) = 0.10$, $p = 0.75$; *novelty*×*continuum*: $F(1, 128) = 9.37$, $p = 0.003$; *novelty*×*group*: $F(1, 128) = 10.87$, $p = 0.001$; *continuum*×*group*: $F(1, 128) = 10.25$, $p = 0.002$). Most important for our purpose is the *novelty*×*continuum* interaction, which demonstrates that the difference between the continua which we have found is a reliable difference. The robust model produced comparable coefficients, suggesting that we can trust the outcome of the least-squares ANOVA.

11.2.3 Conclusions

The analysis has demonstrated that with the OLD continuum – the continuum known from the training – the PHONEMIC group has indeed performed in a more categorical manner; as

predicted by both types of word recognition models. This is an indication that the phonetic categorisation task behaved as we expected it to.

The results for the two NEW continua indicate that the difference between groups can transfer to stimuli not encountered in the training, as predicted by *mediated-access* models; but only in the case where the position of the crucial phonetic contrast remains the same as in the training. A three-way ANOVA showed that the difference between the two types of continua is reliable. I will interpret and further discuss this outcome in §12.3.

Part III

Discussion and conclusions

This last part contains a discussion of the results of my experimental study in Chapter 12, both on their own terms and in light of other findings. This followed in Chapter13 by a summary of the major findings of this thesis, and the conclusions that can be drawn from them.

12 / Discussion of the experimental results

In this discussion chapter I will begin by reviewing the results of the experiment (§12.1). I will then try to resolve the apparent conflict between the outcome of the repetition priming and phonetic categorisation test (§12.2), concluding that the repetition priming task may not have worked as planned and that we should therefore base our conclusions mainly on the phonetic categorisation test. Then, I will interpret the outcome of the phonetic categorisation task (§12.3) and compare my conclusion with the findings of other, similar studies (§12.4). My main conclusion will be that *mediated-access* models are better supported by the evidence. In addition, a comparison of the two different acoustic continua suggests that the prelexical representations used for auditory word recognition are likely to be *position-specific segmental* representations – either in the form of position-specific allophones, the preferred interpretation, or positionally restricted phonemes – and definitely not larger sublexical units such as syllables or syllable codas. I will conclude the chapter by making suggestions for further experiments (§12.5).

12.1 Review of the experiments

The experiment had two parts: a training phase and a test phase. In the two sessions of the training phase subjects were taught to distinguish two minimal pairs. For the PHONEMIC training group, these pairs spanned the /f–ɸ/ contrast; for the ALLOPHONIC group the training contrast was /f–θ/, and [f] and [ɸ] were presented as free variants of the phoneme /f/. Two test were then performed on these two groups; test for which *direct-* and *mediated-access* models make different predictions.

In the *repetition priming* test, subjects had to make lexical decisions on auditory stimuli. Stimuli that are repeated are generally processed faster than stimuli which occur for the first time. This kind of facilitation was observed in my experiment too. Probes that were repetitions of earlier occurrences (the IDENTICAL condition) were processed faster than probes that had

primes which differed by one phoneme (the UNRELATED condition); for example, [nəʊf] was responded to faster when subjects had heard [nəʊf] before than when they had heard [nəʊk].

This outcome is entirely consistent with the literature, and was predicted by both models of word recognition. It was with regard to the RELATED priming condition – where stimuli ending in [f] were paired with stimuli ending in [ɸ] – that *direct-* and *mediated-access* models made different predictions. *Mediated-access* models predict that the PHONEMIC group should treat [nəʊɸ] as different from [nəʊf], and that there should consequently be no facilitation. The ALLOPHONIC group, on the other hand, should treat [nəʊf] and [nəʊɸ] as identical, since for them both stimuli contain the same phoneme /f/; we thus expect a similar amount of facilitation for the RELATED condition as for genuine repetitions in the IDENTICAL condition. *Direct-access* models predict no such interaction of the variables *priming relationship* and *training group*: the two training groups should produce a comparable amount of facilitation in the RELATED priming condition.

The repetition priming task did not result in the interaction predicted by *mediated-access* models. Subjects in the PHONEMIC training group did not perform any differently with RELATED prime-probe pairs than subjects in the ALLOPHONIC training group. This is consistent with DIRECT-ACCESS models. However, there is also some indication that the experiment has not worked quite as expected (see §10.3). This will be further discussed in §12.2 below.

In the *phonetic categorisation* test, subjects had to categorise stimuli of two [f–ɸ] continua: one which was based on a training pair [pəˈkif–pəˈkiɸ], and an entirely new one which subjects had never heard before (either the *position* continuum [ˈfelət–ˈɸelət], or the *vowel* continuum [sɑf–sɑɸ]). *Mediated-access* models predict that, because subjects in the PHONEMIC group have acquired sublexical representations for [ɸ] during the training, they can use these representations in the categorisation task, and this will lead to a more categorical performance with both the OLD and the NEW continuum, compared to the performance of the ALLOPHONIC group. *Direct-access* models also predict a more categorical performance of the PHONEMIC group with the OLD continuum, but they do not predict a transfer of this more categorical behaviour to the NEW continuum.

The results of the test support the MEDIATED-ACCESS account of word recognition: the more categorical performance of the PHONEMIC group can indeed transfer to a completely novel continuum. There appear to be strong constraints on when a transfer to a novel continuum will occur. The PHONEMIC group performed more categorically when the stimulus-final position of the [f–ɸ] contrast of the training was retained in the test (the [sɑf–sɑɸ] continuum), but not when the same contrast occurred in initial position (the [ˈfelət–ˈɸelət] continuum). A straightforward explanation of this difference would be to claim what is acquired in the training is not

a new *phonemic* representation for the sound /ɸ/, but a *position-specific allophonic* representation. What subjects have acquired in the training is a representation for [ɸ] in syllable- or word-final position, that cannot be used to process words where [ɸ] occurs in initial position. This interpretation of the data will be further discussed in §12.3.

For the moment, the main conclusion is the realisation that the outcome of the two tests are in conflict. Repetition priming appears consistent with *direct-access* models, while phonetic categorisation seems to favour *mediated-access* models. How can this apparent conflict be resolved?

12.2 Ways of resolving the conflict between the tasks

There are two ways in which conflicting outcomes that result from different test tasks may be resolved. The first may be called a *substantive* resolution; it consists in suggesting that one of the tests used is more appropriate or reliable than the other. The response variable of psycholinguistic experiments are behaviour measures (unless we are doing brain imaging), and we use them to test theories about unobservable cognitive functions or processes. Consequently, there is always a gap between an experimental result and its substantive interpretation. Some experiments reflect cognitive functions more directly than others. The claim in our case would be that repetition priming is a better test for the existence of prelexical representations than phonetic categorisation.

The second solution would be to remember that statistical hypothesis tests are asymmetric. There is a difference between data that is consistent with a predicted effect (as is the case for the phonetic categorisation test) and data that is consistent with the absence of a predicted effect (as is the case for the repetition priming data). I call this the *statistical* solution.

A third possibility I should briefly mention would be to claim that the conflict need not be resolved. This would mean that the two test tasks together provided evidence that lexical access can be both direct and mediated. My experiment would then speak for a hybrid account, that would acknowledge a prelexical level of processing but also allow this level to be bypassed. In this case, we would still have to conclude, however, that the two test tasks are different; and the issue whether one may provide a more accurate picture of what is going on in auditory word recognition still needs to be addressed.

12.2.1 Substantive resolution

Repetition priming is an *online* task, while phonetic categorisation clearly is not. The term 'online' is used to refer to tasks that measure a cognitive process as that process is unfolding

or, more generally, a task that engages the process that we study. The lexical decision task used in repetition priming is an online task with regard to word recognition: in order to accept a stimulus as a word or reject it as a nonword, subjects have to engage in word recognition. Reaction times can thus be taken to reflect the difficulty that any given type of trial poses to the subject.

Phonetic categorisation is not a task that engages word recognition. Judging which phonetic category a given sound belongs to or, in the case of the AXB-task used in my experiment, judging the similarity of that sound to two other sounds, can be done without the necessity to access lexical representations. In addition, the main response variable of a phonetic categorisation task is not reaction time, but the judgement that the subject makes. This is characteristic of *metalinguistic* tasks. Repetition priming thus seems to reflect the process of word recognition more directly than phonetic categorisation.[1]

In addition to the disadvantage that phonetic categorisation is not an online task with respect to auditory word recognition, one could argue that phonetic categorisation creates the very representations we are trying to test for. The argument is as follows. Because in my phonetic categorisation task I only vary one segment – from a clear [f] to a clear [ɸ] – subjects may form an ad hoc representation for that segment as they perform the task. So rather than providing evidence for the acquisition of segmental representations in the training, the phonetic categorisation task may create these representations.

It is hard to say whether segmental representations can be formed 'on the fly', as suggested by this argument. But even if we accept it, the difference between the two training groups still needs to be explained. No matter what the drawbacks of the phonetic categorisation task are, this difference between the training groups can only be due to differences in training – unless we want to make the highly unlikely claim that the difference is simply due to variation between subjects. Even if segmental representations *are* generated by the phonetic categorisation task, we have to explain why we find more evidence for the existence of segmental representations for the PHONEMIC group than we do for the ALLOPHONIC group.

12.2.2 Statistical resolution

Null hypothesis significance tests are asymmetrical, as we all know. What the *p*-value of the test statistics tells us is the probability of the observed data assuming that the null hypothesis is true. A low *p*-value means that the data is unlikely to have occurred under the null hypothesis;

[1] Notice that, strictly speaking, both tasks are metalinguistic: the lexical decision task also requires subjects to make linguistic judgements. The crucial differences, however, are that (i) the response variable is not the judgement itself but the time it takes to perform the task, and (ii) that lexical decision requires word recognition while phonetic categorisation does not.

and this is normally taken as evidence in favour of the alternative hypothesis. A failure to reject the null hypothesis can, however, not been taken as evidence in favour of the null hypothesis.

This asymmetry is problematic for my experiment, because in both my experimental tests it was the *mediated-access* model that predicted an effect – a more categorical performance in the phonetic categorisation task, and a *training group*×*priming relationship* interaction in the repetition priming task – and the *direct-access* model merely denied the effect. From the point of view of significance testing, the *direct-access* model thus merely states the null hypothesis. The consequence is that we can say that the phonetic categorisation task supports the *mediated-access* model, because the effect it predicted is unlikely to have occurred under the null hypothesis; but we should be reluctant to say that the repetition priming task supports the *direct-access* model, because in this case we merely failed to reject the null hypothesis.

Fortunately, we have a means of assessing how well the *direct-access* model is supported by the repetition priming data: point and interval estimates. Consider the boxplots for the repetition priming data in FIGURE12.1. These are the same as in FIGURE 10.3, but 'zoomed in' on the boxes and without the whiskers. This highlights the notches of the boxplots, which represent approximate 95% confidence intervals for the median.[2]

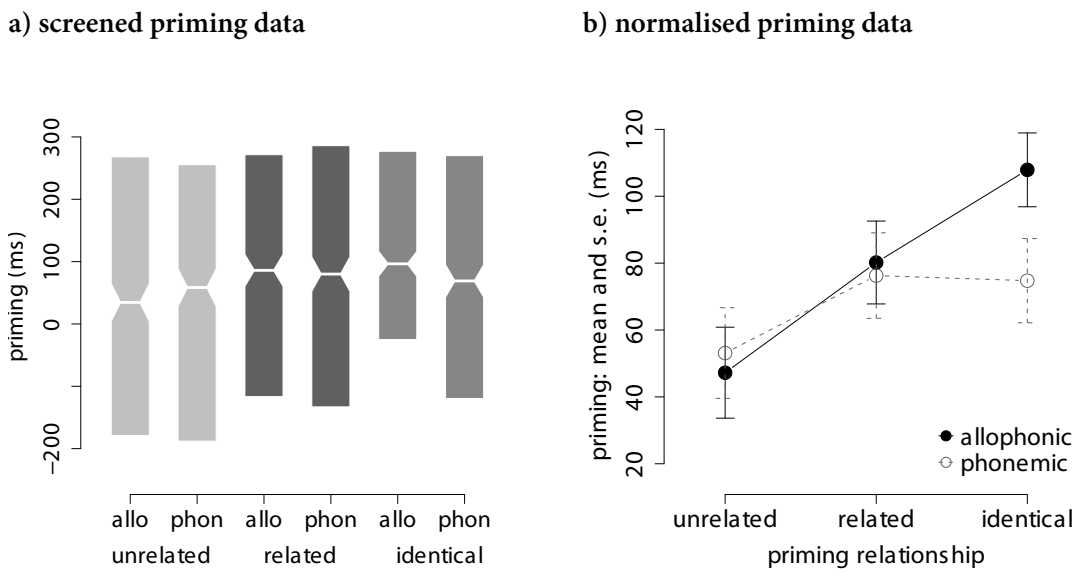


FIGURE 12.1: Boxplots of the screened priming data (left), and interaction plot of the normalised priming data (after removal of outliers). Both plots compare the two *training groups* at the three levels of the *priming relationship*. The boxplots only show the median (horizontal line) and the interquartile range (the box); the notches represent approximate 95% confidence intervals for the median. The interaction plot shows means and standard errors.

[2]The notches are given by $\pm 1.58 IQR/\sqrt{n}$, where *IQR* is the interquartile range, and *n* the sample size (R Development Core Team, 2006, see entry on `boxplot.stats`). Boxplots with notches have been proposed by McGill et al. (1978).

The 95% confidence intervals are quite large – between 40 and 61 ms – but not unusually so; Pallier et al. (2001) present confidence intervals (for means) of about 30 to 80 ms. It is evident though that the two groups hardly differ in the RELATED condition; the two medians are 86 ms (ALLOPHONIC) and 80 ms (PHONEMIC). The failure to find evidence supporting *mediated-access* models does thus not appear to be a problem of lack of statistical power; even with much smaller confidence intervals would this difference of only 6 ms fail to be significant. We may conclude that there is no *training group* difference in the RELATED *priming condition*. So at least in this respect, the repetition priming data can be said to support *direct-access* models. But, as discussed in §10.3, there are reasons to suspect that the repetition priming experiment did not work as expected, and that the results may therefore not provide evidence relevant to the research question. There are two major problems.

First, it appears that there could be a between-group difference in the IDENTICAL condition. This is very clear in the normalised data set (see the interaction plot in FIGURE 12.1), but it is also evident in the non-normalised data if comparing medians (see the boxplots in FIGURE 12.1). This difference is not significant at the 5%- level, but with $p = .10$ it is unlikely enough to arouse suspicion. If there is indeed a difference between the training groups in the IDENTICAL condition, then one of the fundamental assumptions of the repetition priming test would be violated. In addition we would require an explanation why the training may have caused such a group difference.

A possible explanation would be that the training has taught subjects in the PHONEMIC group to pay more attention to the small differences between [f] and [ϕ], and that this results in less priming in the IDENTICAL condition. This is very speculative, and it does also not explain why the same does not apply to the RELATED condition (see again the interaction plot in FIGURE 12.1). Further support comes from an analysis that compared reaction times to primes and probes separately. We found that the PHONEMIC group responded more slowly to both primes and probes in the RELATED condition (and to some extend also in the IDENTICAL condition) than the ALLOPHONIC group. This would also seem to suggest that the PHONEMIC group was slowed down by their paying more attention to small differences.

Whatever the explanation of these differences, they make it difficult to interpret the repetition priming data, and this in turn hampers the direct comparison of the groups in the RELATED condition. In the end, it is probably best to withhold judgement, and to conclude that the repetition priming test was not conclusive.

12.3 Interpretation of the phonetic categorisation data

If we accept that the repetition priming test was inconclusive, we have to focus on the phonetic categorisation test, and ask what it can tell us about the role of sublexical representations in auditory word recognition.

We have seen that the more categorical performance of the PHONEMIC group with the OLD [pəˈkif–pəˈkiɸ] continuum transferred to one of the two NEW continua: the [sɑf–sɑɸ] continuum, where the word-final position of the [f–ɸ] contrast was maintained but a different vowel preceded it. In the case of the [ˈfelət–ˈɸelət] continuum, where the position changed to word-initial, the PHONEMIC training group did perform no more categorically than the ALLOPHONIC group. Can we conclude from this that subjects in the PHONEMIC group have acquired prelexical segmental representations for [ɸ] while subjects in the ALLOPHONIC group have not – as predicted by a mediated-access model?

The acquisition of prelexical segmental representation would explain the outcome. But as with all behavioural experiments performed to gain insight into cognitive functions, there is a gap between theory and data, and several possibilities how that gap may be bridged. There are at least four issues we need to consider:

1) The extension or size of the category acquired: is it a segment, or could it be some other unit?
2) The abstractness of the category: is it a phonological category, or maybe a less abstract representation?
3) The locus of the category: is it a pre- or a postlexical representation?
4) The strength of the inference from phonetic categorisation to cognitive categories.

12.3.1 Are the new representations segments?

The NEW [sɑf–sɑɸ] continuum has only one thing in common with the training stimuli and consequently the [pəˈkif–pəˈkiɸ] continuum: its final segment. The more categorical performance of the PHONEMIC group can be explained by assuming that, in the training, subjects in the PHONEMIC group have acquired a new representation for the segment [ɸ], which enabled them to keep [sɑf] distinct from [sɑɸ] in the phonetic categorisation task. Subjects in the ALLOPHONIC group, however, did not; thus their less categorical performance.

If the PHONEMIC group acquired a new segmental representation, why could they not use it with the [ˈfelət–ˈɸelət] continuum, and perform in a more categorical manner too? This could

be because in the training the fricatives [f] and [ɸ] appeared in the rhyme of the syllable, but in the [ˈfelət–ˈɸelət] continuum in syllable onset. This raises the question whether rhymes and not segments have been acquired. Position (rhyme vs. onset) indeed seems to be crucial, but the units acquired cannot be rhymes because the rhymes used in the training were [if], [iɸ] and [ef], [eɸ], and the rhymes used in the phonetic categorisation task were [ɑf] and [ɑɸ]. The acquisition of whole rhymes thus cannot explain the data for the [sɑf–sɑɸ] continuum.

In both the training and the phonetic categorisation materials, the final segment was also the syllable coda (i.e. the rhyme minus its nucleus), which makes it conceivable that subjects in the PHONEMIC group have acquired codas and not segments. The segment is, however, a simpler unit than the coda, as codas depend on syllable rhymes and onsets. Thus by the principle of parsimony, we have to conclude that what has been acquired were segments and not codas – unless we uncovered additional evidence that could not be explained by segmental representations but required the acquisition of syllabic constituents.[3]

12.3.2 Are the new representations phonemic?

If the newly acquired representations are segments, the default assumption – again following from the principle of parsimony – is that they are phonemes. The concept of *phoneme* expresses the intuition that a speech sound is the same, i.e. belongs to the same category, regardless of where in a word or syllable and in what segmental context it occurs.

The finding that the PHONEMIC group performs more categorical with the [sɑf–sɑɸ] continuum but not the [ˈfelət–ˈɸelət] continuum may be taken to mean that the segments acquired in the training task are not phonemes. Had the PHONEMIC group acquired phonemes, the argument goes, the difference in position between the training (where the contrast occurred in word- or syllable-final position) and the [ˈfelət–ˈɸelət] continuum (word- or syllable-initial position) should not matter, and there should be a difference between training groups for this continuum too.

The most important conclusion to be drawn from this failure of the training to have an effect on the [ˈfelət–ˈɸelət] continuum is that positional information somehow has to be included in the mental representation of the segment. This can be achieved straightforwardly by assuming that the representations acquired are positional allophones, which may be represented by either $[ɸ]_{coda}$ or $[ɸ]_{wordfinal}$, where the subscript specifies in what position the sound was

[3] I have not discussed features because features are generally thought of not as units that have extension but as properties of segments: distinctive features distinguish phonemes. The feature $\pm voice$, for example, distinguishes voiced phonemes from unvoiced phonemes (e.g. /p/ from /b/); and it can be realised by the acoustic cues voice onset time, stop closure duration, stop closure voicing, duration of the preceding vowel, etc. Nonetheless, a featural account would be possible, as long as we made sure that the features are associated with segmental representations and not any larger units.

encountered and acquired. Such position-specific representations would explain why generalisation from the training pair [pəˈkif–pəˈkiɸ] to the [ˈfelə[…]

There is an alternative interpretation, however; namely that the representations acquired are positionally restricted but nonetheless *phonemic* representations: /ɸ/$_{coda}$ instead of [ɸ]$_{coda}$, in other words. This interpretation is made possible by the fact that many languages have speech sounds which are strongly restricted with regard to their syllabic position, but which are still best regarded as phonemes. Examples from English are /ŋ/ and /h/: the former only occurs in syllable rhymes and the latter only in syllable onsets. The main reason why they are not regarded as conditional allophones of the same phoneme – unlike clear and dark /l/ ([l] and [ɬ]), which also occur in complementary distribution – is that /ŋ/ and /h/ are very different sounds: /ŋ/ is a nasal with a velar place of articulation and /h/ a glottal fricative, whereas [ɬ] differs from [l] only in having a secondary articulation.

The issue whether phonemic representations with an optional positional restrictions, /ɸ/$_{coda}$, or positional allophones, [ɸ]$_{coda}$, have been acquired in the training is clear enough from a phonological point of view, but it is not obvious whether this difference is amenable to psycholinguistic testing. In order to test whether a contrast transfers to positions in which it was not encountered in the training, its distribution in the training obviously needs to be restricted to a limited number of positions. If the training contrast does not transfer – as was the case in the present study – the phonological issue of phoneme vs. allophone is unanswerable. Only if a transfer does occur can we give an unambiguous answer, because in this case we know that there is no positional restriction and that the speech sound has, therefore, been acquired as a phoneme. I address this problem again in §12.5 when discussing additional experiments.

12.3.3 Are the new representations prelexical?

We have learnt so far that the representations which the PHONEMIC group has acquired in the training are likely to be segmental representations, and that they need to be less abstract than phonemes. In all likelihood, they are positional allophones. The next question is whether they are also *prelexical* representations.

The phonetic categorisation task itself does not tell us anything about the locus where the categories are acquired. Because phonetic categorisation is a metalinguistic task which does not depend on word recognition, we cannot say whether any of the categories it provides evidence for are prelexical or postlexical. In my experiment, however, the phonetic categorisation task has been performed after two training sessions, and the training task did required subjects to recognise words. I would like to argue that if the acquisition of words caused the acquisition of

segmental representations, these segmental representations are prelexical, i.e. they have been formed because they are *required* for auditory word recognition.

While I think that this argument is valid in principle, it is weakened by the nature and number of the training stimuli. The training task arguably requires subjects to recognise new words; but the training stimuli were just two minimal pairs. Because of the small number of training stimuli and because they were minimal pairs, subjects attention was likely to have been drawn to the stimulus-final contrast.[4] As a result of this, we cannot exclude the possibility that the representations acquired are postlexical and not prelexical.

There are thus arguments both ways. The fact that the training required subjects to recognise words suggests that the sublexical representations thereby generated should be prelexical. The fact that the training involved only a small set of words that were, moreover, minimal pairs makes it possible that only postlexical representations have been acquired. In consequence, we cannot make any firm conclusion about the locus of the sublexical representations acquired. Evidence from similar studies (particularly McQueen et al., 2006; see §12.4) make a prelexical locus seem more likely, however.

12.3.4 The phonetic categorisation task as evidence for categories

The final question I want to address is how good the evidence from phonetic categorisation is as demonstration of existence of mental categories. Take a look again at the idealised predictions of *mediated-access* models and the actual outcome of the phonetic categorisation task (both repeated in FIGURE 12.2).

It is obvious that the experimental results are nowhere near as clear-cut as the idealised predictions. In the case of the OLD continuum, even the responses of the ALLOPHONIC training group appear quite categorical. The reason for this might be (i) that subjects already had a category for the [f]-end of the continuum; and (ii) that exposure to the end points of a continuum, even if they are presented as the same phoneme, can result in a categorical performance with an AXB task. It is equally obvious, however, that there is a difference in the degree of categoriality between the two training groups (and that it is a significant difference we have seen in §11.2). The same is true for the NEW continuum: the PHONEMIC group responded in a more categorical way than the ALLOPHONIC group. With this continuum, the categorisation function of the ALLOPHONIC group is distinctly continuous; but that of the PHONEMIC group, while noticeably more S-shaped, is far from an ideal categorical response.

That experimental evidence is gradient and relative is not unusual. In our case, it is the differ-

[4] Note, however, that participants were in general unable to put into words what the difference was between minimal pairs; most seemed to think that the difference was durational.

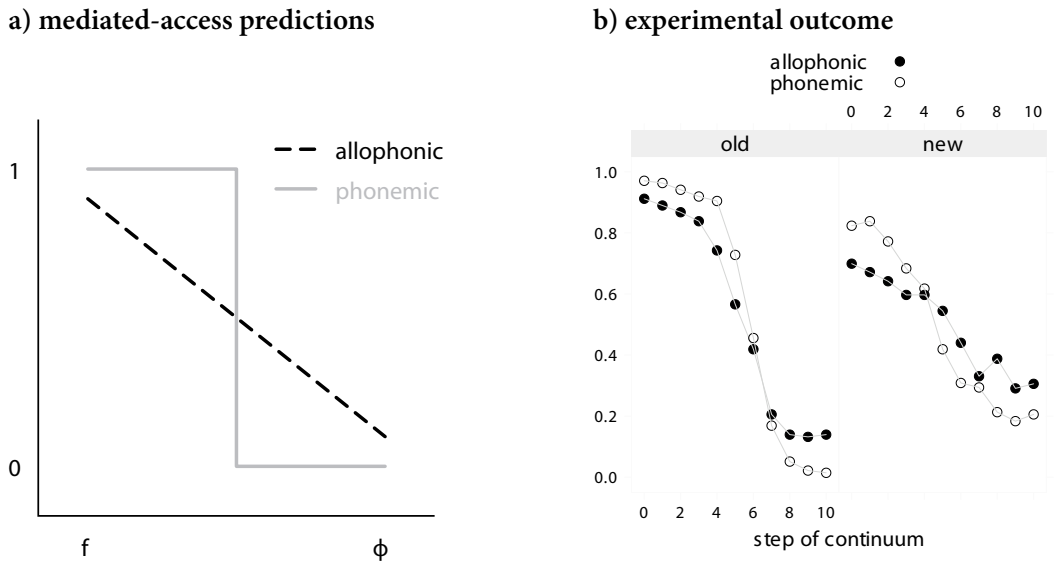


FIGURE 12.2: Ideal and reality. Idealised predictions of *direct-access* models (left) and experimental outcome for OLD (pəˈkif–pəˈkiɸ) and NEW (sɑf–sɑɸ) continua (right). See running text for discussion.

ence between the training groups that counts, and not some absolute measure of categoriality. But it would be a good thing if the evidence for or against cognitive categories were itself more categorical. We might ask whether the difference between having a segmental representation for the sound [ɸ] and not having one, should not be larger than a small (although clear and consistent) increase in categoriality. A test which provided an absolute measure of the existence of phonological categories would be welcome; but as far as I know, there is none available. The Haskins conception of categorical perception (see §6.3) which defined it in terms of a predictive relationship between a phonetic categorisation and discrimination task, could be regarded as an attempt to provide an absolute test for phonological categories. But as we have seen, categorical perception cannot be used in this way, because the link between categorisation and discrimination is tenuous, and because not every phonological category is perceived categorically in the Haskins sense.

Short of a better test for the existence of segmental categories, and bearing in mind that the subjects in my experiments underwent only little training, it bears repeating that the results of the AXB task have been quite clear. We can therefore say that the phonetic categorisation test provides relatively good evidence that the PHONEMIC training group has indeed acquired a new representation for [ɸ], while the ALLOPHONIC group has not. This is consistent with *mediated-access* models.

12.4 Comparison with other studies

The results of my experimental study, particularly the phonetic categorisation task, suggest that *sublexical* representations are used for auditory word recognition. These representations are likely to be *segmental*, and need to be less abstract than phonemes: probably some form of *positional allophones*. In addition there is some evidence that these representations are *prelexical*, in the sense of being required for auditory word recognition; though we cannot make any strong claims about their locus. In this section I will compare my findings with those of other studies.

12.4.1 Prelexical representations

We have seen in Chapter 4 (§4.3 and §4.4) that there are two studies that have presented strong evidence for a prelexical locus of sublexical representations: Pallier et al. (2001) and McQueen et al. (2006).

Pallier et al. (2001) found that Spanish-dominant Catalan-Spanish bilinguals produced an equal amount of facilitation in a repetition priming task with Catalan minimal pairs (e.g. /osos/ 'bear' vs. /ɔsos/ 'bones') as they did with identical pairs; Catalan-dominant bilinguals, on the other hand, show no priming with Catalan minimal pairs. We can conclude from this that Catalan minimal pairs are treated as homophones by Spanish-dominant bilinguals.

A *mediated-access* model with prelexical segmental representations can account for these findings, if we assume that the Spanish-dominant bilinguals use their Spanish prelexical representations to process Catalan words, while the Catalan-dominant bilinguals have separate prelexical representations for the phonemes that are specific to Catalan. A *direct-access* model, which claims that listeners store only whole words, has no easy way of accounting for the categorical difference in the performance of the two groups. It is not obvious, however, whether we can draw conclusions to other populations, such as less proficient bilinguals, bilinguals speaking different languages, or monolingual speakers.

McQueen et al. (2006) found that a lexical bias acquired for a fricative that is ambiguous between [f] and [s] generalises to words that have never before been encountered with the ambiguous fricative. In a cross-modal priming task, auditory stimuli containing the ambiguous fricative facilitated the processing of corresponding visual stimuli containing letter 'f' only for subjects who had previously learnt to treat the ambiguous fricative as an /f/, but not for those who had acquired an /s/-bias. The locus of this bias is likely to be prelexical, because it has a direct influence on lexical decisions, and because it transfers to new stimuli that only have the ambiguous fricative in common with the training stimuli.

The results of my phonetic categorisation task do not provide indisputable evidence for the existence of prelexical representations; but they are clearly consistent with *mediated-access* models. In addition, the results are consistent with McQueen et al.'s (2006) findings in as much as both studies show that lexical learning can have an influence on sublexical processing. Unfortunately in my case, the repetition priming test, which could have provided stronger evidence for the *prelexical* locus of the process, did not work as expected.

12.4.2 Segmental representations

The studies by Pallier et al. (2001) and McQueen et al. (2006) are generally consistent with my finding that the prelexical representations are segmental. Segment-sized prelexical representations can account for the Spanish-dominant bilinguals' treatment of Catalan minimal pairs as homophones, as well as McQueen et al.'s finding that the lexical bias acquired for an ambiguous fricative extends to novel stimuli that contain the same ambiguous fricative.

Direct evidence in support of the segment has been presented by Eisner and McQueen (2005). They used the same perceptual learning paradigm as McQueen et al. (2006), but with phonetic categorisation as their test task. They studied the effect that a change of speaker identity has in the perceptual learning paradigm. The lexical bias acquired in the training – i.e. whether a subject regards the ambiguous fricative as an /s/ or /f/ – resulted in a boundary shift in the phonetic categorisation task when the fricatives used in the training and the test were produced by the same speaker; when they came from different speakers, there was no boundary shift in the phonetic categorisation task. Whether or not the speaker of some other part of the stimulus changed, had no effect on the phonetic categorisation performance.

These findings agree with the results of my phonetic categorisation test with the [sɑf–sɑɸ] NEW continuum. As long as the segment used in the training – the ambiguous fricative in Eisner and McQueen's study and the /f–ɸ/ contrast in my study – recurs in the test task, we get an effect; the immediate context of the segment is allowed to change. This suggests that, in both studies, the training was constrained to segmental representations. In the perceptual learning paradigm, the segment is where the adjustment seems to take place, and in my experiment, it is where the learning occurs.

12.4.3 Position-specific representations

Pallier et al.'s (2001) study does not allow us to distinguish phonemes from more concrete representations, such as position-specific allophone. Priming between e.g. /osos/ and /ɔsos/ shows that Spanish-dominant bilinguals treat the Catalan sounds [o] and [ɔ] as being identical. This

means that they have only one representation for both [o] and [ɔ], but whether this representation is position-specific or not we cannot say.

All studies that have used the perceptual learning paradigm with the ambiguous fricative sound (at least all the ones reported so far: Norris et al., 2003, Eisner and McQueen, 2005, McQueen et al., 2006) have presented it in the same stimulus-final position in both the training and the test. These studies do also not allow us to address the issue of whether representations are position-specific.

McQueen and Mitterer (2005) used ambiguous vowels, and presented the vowels in the training phase as the nuclei of the final syllable of polysyllabic words, and in the test phase in [Vft] continua, where the V stands for the ambiguous vowels. They did find the usual boundary shift in the phonetic categorisation test. Note that in this experiment, the numbers of syllables changed between training and test, and the consonants that followed the ambiguous vowels. This is thus comparable to my [sɑf–sɑɸ] continuum. McQueen and Mitterer did not study a case where the position of the ambiguous vowel changes between training and test (corresponding to my [ˈfelət–ˈɸelət] continuum). There is thus again no way of telling whether the representations where the adjustment takes place are position-specific representations.

A series of experiments that found results comparable to mine was reported by Dahan and Mead (2006). Dahan and Mead extended a perceptual learning experiment with noise-vocoded speech by Davis et al. (2005). Noise-vocoded speech is created by dividing a speech signal into frequency bands, extracting a smoothed amplitude envelope for each band, and applying this envelope to white noise; thereby creating a signal which retains most of the temporal characteristics of the original speech signal but loses most of its spectral characteristics.[5] Noise-vocoded speech is unintelligible at first, but Davis et al. (2005) showed that subjects can quickly adapt to it, particularly if they are presented with the same sentences produced in clear and distorted speech.

Dahan and Mead (2006) used noise-vocoded speech to see how fine-grained the adjustment to this type of speech is. In one of their experiments they presented subjects with CVC stimuli both distorted and in the clear. The first and second consonant came from a small set: [f] and [d] in initial position and [m] and [t] in final position. Subjects were later tested on their identification of consonants in distorted stimuli. Two of the consonants occurred in the same position as in the training, e.g. 'fan' [**f**an] and 'nut' [nʌ**t**]; the two other consonants occurred in a different position, e.g. 'mall' [**m**ɔl] and 'loud' [laʊ**d**]; a third set of test words contained consonants not encountered in the training phase. Dahan and Mead showed that only if the consonants from the training were presented in the *same syllabic position* was their identifica-

[5] See http://www.mrc-cbu.cam.ac.uk/∼matt.davis/vocode/ for an example.

tion significantly better than the identification of consonants that have not been heard in the training; when the position changed between training and test, subjects' performance was no better than with the new consonants. These findings are consistent with our conclusion that segmental representations are position-specific; but because Dahan and Mead (2006) do not introduce new speech sounds, the further issue of whether the representations are positional allophones or positionally restricted phonemes is not touched by it.

Additional evidence that prelexical representations, if they exist, are likely to be position-specific has been reported by Ganong et al. (2001). They used a form priming paradigm with several test tasks, and compared prime-target pairs with different kinds of overlap; for example, /bæt/ could be primed by itself, or by /but/ (different vowel), or by /tæb/ (the same phonemes in the reverse order), etc. Ganong et al. used these prime-target pairs to compare three types of representations: *phonemes, position-specific* representations (which keep track of their position in the syllable), and *context-sensitive* representations (which not only keep track of their position but also of the neighbouring representations). How many overlapping segments there are depends on which type of representation is used: in the case of /tæb/ and /bæt/, three phonemes are shared, but only one position-specific representations (the /æ/), and no context-sensitive representation. If the amount of priming depends on the amount of shared segmental representations, then these different representations make different predictions about how much /tæb/ primes /bæt/. The pattern of priming that Ganong et al. (2001) found was most consistent with the predictions of position-specific representations.

Neither Dahan and Mead's (2006) study nor Ganong et al.'s (2001) study directly compared *direct-* and *mediated-access* models; but their findings are consistent with my finding that the representations acquired in the training task have to be position-specific.

12.5 Prospects

In this final section, I will consider ways to extend and improve the existing experimental paradigm.

12.5.1 Expanding phonetic categorisation

The phonetic categorisation test can be easily modified by using different types of continua. There is the additional advantage that if an experiment is run with phonetic categorisation as the only test task, we will have greater freedom to make modifications to the training task (see the next section).

In the current experiment, the following stimuli have been used (for the PHONEMIC training

group):

Training pairs	[pəˈkif–pəˈkiɸ]	[tɪnˈdef–tɪnˈdeɸ]
Test continua	[ˈfelət–ˈɸelət]	[sɑf–sɑɸ]

Between training and test there was thus a change in position (ˈfelət–ˈɸelət), or immediate segmental context and numbers of syllables (sɑf–sɑɸ). These variable allow for further modifications.

First, we have seen that a change of the vocalic context from front vowel in the training to back vowel in the test does not block the use of the newly acquired sublexical category. We would expect this to hold for all vowels. This can be tested e.g. by using a rounded vowel in the test, for example the continuum [sɔf–sɔɸ].

We have also learnt that the effect of the training is position-specific; but we could not determined whether the domain is the word or the syllable. This could be discovered by using two sets of continua with word-medial stimuli, e.g. the following:

Training pairs	[pəˈkif–pəˈkiɸ]	[tɪnˈdef–tɪnˈdeɸ]
Test continua	[mifˈten–miɸˈten]	[miˈfen–miˈɸen]

If it is position within the word that matters both continua should be treated alike. Given that the training in final position did not transfer to initial position, we would expect it not to transfer to medial position either; but whatever will happen, it ought to affect both continua equally. If it is syllabic position that matters, we would expect a training group difference for the [mifˈten–miɸˈten] but not for the [miˈfen–miˈɸen], because in the first continuum has the crucial [f–ɸ] contrast in the same syllable-final position as in the training, while in the second continuum the contrast is in syllable-initial position.

In a similar way we could test for whether subjects acquire segments or rather features in the training. Take for example the following training pairs and test continua:

Training pairs	[pəˈkif–pəˈkiɸ]	[tɪnˈdef–tɪnˈdeɸ]
Test continua	[pəˈkif–pəˈkiɸ]	[pəˈkiɱ–pəˈkim]

(The symbol [ɱ] stands for a labiodental nasal.) If an abstract place feature is acquired in the training, the more categorical performance of the PHONEMIC training group should transfer to the nasal continuum. Such a transfer seems rather unlikely, and I would expect this experiment to come out in favour of segments.

Extending the phonetic categorisation task does, unfortunately, not help us to directly distinguishing between positional allophones and positionally restricted phonemes (as discussed in §12.3); introducing a contrast with a positional restriction always allows both interpretations. Solace may be found in numbers, however. If several different phonetic categorisation

tasks are carried out to test whether the training contrast transfers to different positions, and we never find such a transfer occurring, it would suggest that *all* segmental representations are indeed position-specific and thus best described as positional allophones. If transfer to a new position occurs for some sounds but not for others, it would, on the other hand, favour positionally restricted phonemes. This strategy requires the use of additional training contrasts, so that enough independent tests are possible. Which brings us to the next topic.

12.5.2 Changes to the training task

The training task can also be modified in many ways. An obvious modification is the use of a larger set of training stimuli. Depending on how large the set is, this modification may require additional training sessions.

The use of more training stimuli would also allow us to introduce more variation into the training task. The training contrast could occur in more segmental contexts, or in different positions (e.g. syllable-initially as well as syllable-finally). In the current experiment, there was only one token each of [f] and [ɸ] that was used in all training stimuli. The training stimuli could come from different speakers, or an acoustic continuum could be used, with both prototypical [f]'s and [ɸ]'s and more ambiguous tokens.

It is a natural assumption that variation in the data encourages abstraction; introducing more variation into the training task may therefore produce stronger evidence for *mediated-access* models. On the other hand, too much variation may make it harder to abstract the recurring features of a stimulus type, so that subjects may resort to storing lexical traces of all encounters. In this case we would expect the predictions of *direct-access* models to be confirmed. In short, the systematic introduction of variation into the training task would allow us to study the condition in which prelexical segmental representations are being formed.

As already mentioned, if we focus on phonetic categorisation as our only test task then changes to the training task become easier to implement. The reason is that the repetition priming task places high demands on the selection of stimuli, and these demands have ramifications for the training task too. For example, the reason why the [f–ɸ] contrast occurred in final position in the training was that it had to occur stimulus-finally in the repetition priming task (see §7.4). Likewise, the training stimuli were introduced as unusual English words used in Sri Lanka, because English words were needed for the lexical decision task (see §7.3.1). The phonetic categorisation task is much more flexible, especially if a categorical AXB task is used. Almost any continuum can be used with an AXB task, and consequently phonetic categorisation puts few restrictions on the training task.

One important consequence of this increased flexibility is that the combination of train-

ing procedure and phonetic categorisation task can very easily be extended to languages other than English. It is not necessary to select a new set of stimuli, nor to re-record them. And even if we want to have language-specific stimuli, their construction is straightforward: all that is required are minimal pairs that allow the construction of acoustic continua. This is a distinct advantage of this method over the perceptual learning paradigm of Norris et al. (2003), Eisner and McQueen (2005), McQueen and Mitterer (2005), McQueen et al. (2006). Their training task requires the construction of an ambiguous sound which is perceived as halfway between two phonemes, by the native speakers of the language studied; and this sound also needs to occur in a fairly large set of words. The stimuli used with Dutch subjects in the studies cited above cannot, therefore, be used with other languages; completely new stimulus sets would have to be selected and recorded. The disadvantage of my experiment – if used only with a phonetic categorisation test – is of course that we cannot be certain about the locus of the representations acquired. But it is a very useful paradigm if our goal is to determine how abstract or concrete segmental representations have to be, and what type of information they contain.

12.5.3 Making repetition priming work better

While a combination of training task and the phonetic categorisation task would be useful, it is still worthwhile to consider if we can improve the repetition priming task. The main problem with the current design seems to be that the PHONEMIC training group performed a training task that was harder than that of the ALLOPHONIC group: the PHONEMIC group had to distinguish words with a non-native contrast, whereas the ALLOPHONIC group was given a native contrast with some unusual within-category variation. This difference in training might have disposed the PHONEMIC group to pay more attention to phonetic detail, as we have seen, and this heightened awareness could have caused a performance which made it impossible to compare the two groups.

If this interpretation is correct, the key to making repetition priming work as a test task would be to try and make the training tasks more equal. One way this could be achieved is by turning the variable *priming relationship* into a within-subjects variable. This would mean finding a second phonetic contrast; the [f–ɸ] contrast could then be presented to one training group as a PHONEMIC contrast and as a ALLOPHONIC contrast to the other group, and vice versa for the second contrast; in this way both groups would get some hard and some easy minimal pairs.

Whether this would have the desired effect of making both groups pay equal attention to detail is hard to predict. There could still be a difference on individual items, with each group paying more attention to the items of *their* PHONEMIC group. Apart from this general problem,

it is also difficult to find a suitable second contrast. The [f–ɸ] contrast had several desirable properties. The first is that one of the pair is an English phoneme. This is necessary in order to present the contrast as an allophonic one. The second is that the difference between [f] and [ɸ] is small but noticeable. Had the difference been too obvious, it might be hard to convince the ALLOPHONIC group that the two sounds are really the same phoneme; had the difference been too small, it might have been impossible for the PHONEMIC group to distinguish their minimal pairs. It is obviously difficult to find a second contrast that satisfies all these criteria. A [ʃ–ç] or [ʃ–ʂ] contrast might be feasible. The postalveolar fricative [ʃ] is the native sound; and the palatal or retroflex fricatives may have the desired properties of being similar to [ʃ] while being noticeably different.

12.5.4 Alternative tests

One of the main disadvantage of my choice of test tasks is that for both the phonetic categorisation and the repetition priming task it is the *mediated-access* model that predicted and effect that the *direct-access* model predicted not to occur. If we could be sure that the repetition priming task has worked as expected, this would not have been that much of a problem. But because we have reasons to doubt this, it is hard to say whether the repetition priming task really contradicts the outcome of the phonetic categorisation task, or whether this apparent contradiction is just due to a failure of the task (as I have in effect argued in §12.2).

This issue could be resolved if we had a task for which it was the *direct-access* model that predicted an effect that the *mediated-access* model denied. Finding such a task is difficult, because only the *mediated-access* model predicts a categorical difference between the two training groups. But maybe we can use the fact that according to the *direct-access* model, the amount of priming should depend on the overall similarity between words. This prediction may be tested by introducing additional differences between primes and probes.

In the current repetition priming task I have used RELATED primes such as [flɔf] for probes such as [flɔɸ]. For these the *mediated-* but not the *direct-access* model predicts a difference between the ALLOPHONIC and PHONEMIC training groups. If we introduced a further difference and used [tlɔf] as a prime for [flɔɸ], we could ask how much further reduction in priming the additional mismatch would cause. *Direct-access* models would predict a similar reduction for both training groups, since it is overall difference that matters. *Mediated-access* models would only predict a reduction for the ALLOPHONIC group, because for the PHONEMIC group priming should already be close to zero with [flɔf] as a prime. These predictions obviously assume that the basic repetition priming paradigm is working properly. This means that the changes proposed here cannot be used a an alternative to the current paradigm, but rather as an addition

to it. The main priority should therefore be to try and make the existing repetition priming test work better.

12.5.5 Conclusions

We may conclude this short section about how to extend and improve the existing experimental paradigms by noting that there are two options. The first is to focus on the phonetic categorisation test. This would simplify the experiment considerably, and could still provide important information about the size and abstractness of the sublexical representations used. It would not provide clear evidence, however, about whether these representations are prelexical. To do this, we need to try and improve the repetition priming test, first by making the training task for the two groups more comparable. If this manipulation proves successful, we may then want to introduce additional variation into the repetition priming test.

13 / General conclusions

This short final chapter is a summary of the major conclusions that can be drawn from this thesis.

In order to classify auditory word recognition models with respect to their levels of processing and the types of representation used at each level, I proposed four descriptive dimensions. The *abstract/concrete* dimension concerns the complexity of the representations used, and it was defined in terms of the numbers of unrelated variables required to define a representation of a given type, and the number of values that these variables can take; it is a continuous dimension. The *exemplar/summary* dimension has to do with the relationship between representation *types* (e.g. words) and their *tokens* or *exemplars*; this was defined as a binary dimension, with *exemplar* representations constantly accepting new exemplars and *summary* representations being unitary and closed. The *structured/unstructured* dimension is also binary; *structured* representations are decomposable into smaller parts, while *unstructured* representation are not. The *prelexical* dimension represents the question whether there are levels of processing – which were defined as processing stages where units of a certain size and abstractness are recognised – previous to the lexical level and required for word recognition; this is a discrete dimension and it was treated as a binary one, with segments as the only prelexical recognition units.

The three binary dimensions *exemplar/summary*, *structured/unstructured* and *prelexical* allow for 24 combinations, only six of which make theoretical sense. The main criterion was that representations at the different levels have to be *commensurate* for a model type to be functional. Of the six remaining types, two would require a hybrid model with mixed representations and two access routes to the lexicon. The four non-hybrid types are:

Summary direct-access model: there are *no* prelexical representations; lexical representations are *summary* and *not structured*, and they have to be *concrete*.

Exemplar direct-access model: there are *no* prelexical representations; lexical representations are *exemplar* and *not structured*, and they have to be *concrete*.

Summary mediated-access model: lexical representations are *summary* and *structured*, and can be *abstract*; prelexical representations are *summary* and have to be *concrete*.

Exemplar mediated-access model: lexical representations are *summary* and *structured*, and can be *abstract*; prelexical representations are *exemplar* and have to be *concrete*.

The *abstract/concrete* dimensions also reduces to a binary dimension due to the requirement of representations to be commensurate. Representations either have to be concrete enough to be compared with the acoustic/auditory input or they can be as abstract as the representations at the next-lower level allow. If, for example, phonemes are recognised at a prelexical level of processing, lexical representations can be specified straightforwardly as strings of phonemes.

The following four questions allow us to distinguish experimentally between these theoretically feasible model types:

1) Is there a prelexical level of processing?

2) Are lexical/prelexical representations exemplar or summary?

3) Are lexical representations structured?

4) Are lexical representations concrete or abstract?

The first question is the most central, and therefore also the one most deserving to be studied. To address question 2 we already need to know whether word recognition is direct or mediated. Questions 3 and 4 do not really allow us to distinguish *direct-* from *mediated-access* models, because lexical representations can show signs of being structured or abstract for other reasons than that word recognition is mediated by a prelexical level of processing. Only the first question thus directly compares *direct-* and *mediated-access* models.

In a survey of experimental studies that have addressed the question whether there is a prelexical level of processing, we found two studies that have reported evidence supporting *mediated-access* models: McQueen et al. (2006), with their lexical learning paradigm, and Pallier et al. (2001), using a repetition priming study with a Spanish-Catalan bilingual population. Some studies have presented evidence suggesting that, in some circumstances at least, lexical access may be direct (Marslen-Wilson and Warren, 1994, McLennan et al., 2003, Connine, 2004).

My own experiment also produced seemingly conflicting results. The outcome of the repetition priming test task was more consistent with *direct-access* models: the *training group*×*priming relationship* interaction that *mediated-access* models predict was not observed. But the repetition priming task also produced some unexpected patterns – particularly a potential dif-

ference between the training groups in one of the control priming conditions – which should make us cautious about drawing firm conclusions from the repetition priming task.

The phonetic categorisation task produced results consistent with *direct-access* models. The different training procedures of the two training groups showed up not only in a more categorical performance of the PHONEMIC training group with the [pəˈkif–pəˈkiɸ] continuum (familiar from the training), but also in a more categorical performance with the [sɑf–sɑɸ] NEW continuum. As well as being consistent with *mediate-access* models, this result also suggest that the sublexical units used in such a model are more likely to be segments than larger units. The additional finding that the more categorical performance of the PHONEMIC training group did not transfer to the [ˈfelə[…

A comparison with other studies, particularly McQueen et al. (2006), makes a prelexical locus seem more likely. The conclusion that the prelexical representations are segments is supported by Eisner and McQueen (2005). None of the previous studies that looked at the issue of prelexical representations for auditory word recognition addressed the issue of how specific these representations are; but other studies (Ganong et al., 2001, Dahan and Mead, 2006) have presented results which are consistent with my conclusion that the representations used contain positional information.

Appendices

A/ Stimuli for the repetition priming task

A.1 Test stimuli

A.1.1 Monosyllabic pseudowords (n=30)

The number after each stimulus gives its duration in ms. The lexical frequency for the competitors listed in the last column is the average of CELEX's CobMln and CobSMln values.

Code	f stimulus		ɸ stimulus		stop stimulus		competitors
T01	frɒf	602	frɒɸ	481	frɒp	662	from (3540)
T02	nəʊf	736	nəʊɸ	515	nəʊk	495	know (681), knows (130)
T03	kʊf	480	kʊɸ	429	kʊp	420	could (391)
T04	gəʊf	570	gəʊɸ	478	gəʊk	501	go (366), goes (224)
T05	həʊf	544	həʊɸ	545	həʊt	682	whole (303)
T06	waɪlf	492	waɪlɸ	601	waɪlp	609	while (295), wilde (49)
T07	maɪnf	690	maɪnɸ	607	maɪnk	621	mind (288)
T08	paʊnf	729	paʊnɸ	631	paʊnk	603	pounds (209), pound (56)
T09	lʊf	549	lʊɸ	471	lʊp	433	look (168)
T10	graʊnf	748	graʊnʔ	652	graʊnk	590	ground (117)
T11	kəʊlf	650	kəʊlɸ	593	kəʊlk	585	cold (116), coal (69)
T12	plʌf	483	plʌɸ	507	plʌt	660	plus (113)
T13	sʌf	642	sʌɸ	626	sʌt	718	son (110), sun (106), [suffer (9), subtle (21)]
T14	flɔf	771	flɔɸ	924	flɔk	779	floor (104)
T15	frɑnf	906	frɑnɸ	997	frɑnt	785	France (102)
T16	bænf	663	bænf	611	bænt	618	bank (99), [banter (1), Bantu (2)]
T17	ʃəʊf	728	ʃəʊɸ	744	ʃəʊk	773	show (97), shown (85), [chauffeur (3)]
T18	tʊf	479	tʊɸ	493	tʊp	536	took (89)
T19	nɔɪf	759	nɔɪɸ	611	nɔɪp	617	noise (71)
T20	saɪnf	872	saɪnɸ	765	saɪnk	720	sign (67), signs (32)
T21	ɑmf	726	ɑmɸ	827	ɑmp	538	arms (65), arm (63)
T22	læf	566	læɸ	538	læt	572	lack (61), [Latin (31)]
T23	hɔf	665	hɔɸ	720	hɔp	565	horse (59)
T24	drɔf	704	drɔɸ	650	drɔk	682	drawn (53)
T25	rɒf	584	rɒɸ	565	rɒp	563	rock (51)
T26	stɒf	711	stɒɸ	640	stɒt	793	stock (44)
T27	twaɪf	658	twaɪɸ	674	twaɪp	679	twice (44)
T28	ʤuf	668	ʤuɸ	717	ʤut	651	June (43)
T29	brɔf	720	brɔɸ	591	brɔp	657	brought (39)
T30	kæmf	699	kæmɸ	713	kæmk	658	camp (36)

A.1.2 Disyllabic pseudowords (n=30)

The number after each stimulus gives its duration in ms. The lexical frequency for the competitors listed in the last column is the average of CELEX's CobMln and CobSMln values.

Code	f stimulus		ɸ stimulus		stop stimulus		competitors
T31	əˈbaʊf	788	əˈbaʊɸ	648	əˈbaʊk	662	about (2952)
T32	pəˈhæf	674	pəˈhæɸ	609	pəˈhæt	837	perhaps (632)
T33	bɪˈfɔf	923	bɪˈfɔɸ	1003	bɪˈfɔt	918	before (516)
T34	wɪˈðaʊf	929	wɪˈðaʊɸ	842	wɪˈðaʊk	903	without (449)
T35	təˈwɔf	976	təˈwɔɸ	834	təˈwɔp	896	towards (212), toward (40)
T36	ˈðeəfɔf	1094	ˈðeəfɔɸ	1056	ˈðeəfɔk	1031	therefore (234)
T37	ˈsʌmwʌf	907	ˈsʌmwʌɸ	1011	ˈsʌmwʌk	951	someone (189)
T38	əˈlɒf	726	əˈlɒɸ	704	əˈlɒk	753	along (173)
T39	rɪˈzʌlf	918	rɪˈzʌlɸ	1016	rɪˈzʌlp	932	result (138)
T40	əˈləʊf	888	əˈləʊɸ	857	əˈləʊt	820	alone (122)
T41	səˈpɔf	1065	səˈpɔɸ	1010	səˈpɔk	804	support (110)
T42	aʊtˈsaɪf	1007	aʊtˈsaɪɸ	987	aʊtˈsaɪp	877	outside (107)
T43	əˈpɑf	923	əˈpɑɸ	978	əˈpɑp	775	apart (104)
T44	əˈkaʊnf	965	əˈkaʊnɸ	1020	əˈkaʊnk	751	account (97)
T45	ˈsəʊvɪəf	974	ˈsəʊvɪəɸ	1017	ˈsəʊvɪəp	927	Soviet (97)
T46	bɪˈjɒnf	915	bɪˈjɒnɸ	876	bɪˈjɒnk	778	beyond (95)
T47	ˈjusfʊf	879	ˈjusfʊɸ	989	ˈjusfʊt	925	useful (91)
T48	ɪnˈsaɪf	937	ɪnˈsaɪɸ	860	ɪnˈsaɪk	903	inside (84)
T49	rɪˈpɔf	986	rɪˈpɔɸ	963	rɪˈpɔk	823	report (82)
T50	θruˈaʊf	931	θruˈaʊɸ	852	θruˈaʊp	895	throughout (78)
T51	ˈtrænspɔf	1190	ˈtrænspɔɸ	1114	ˈtrænspɔp	987	transport (72)
T52	ˈkɒntæf	892	ˈkɒntæɸ	864	ˈkɒntæp	961	contact (67)
T53	təˈnaɪf	793	təˈnaɪɸ	851	təˈnaɪk	705	tonight (67)
T54	dɪˈzaɪf	830	dɪˈzaɪɸ	931	dɪˈzaɪt	833	design (54)
T55	əˈtæf	727	əˈtæɸ	720	əˈtæt	807	attack (53)
T56	ˈfʊtbɔf	989	ˈfʊtbɔɸ	823	ˈfʊtbɔt	945	football (52)
T57	bɪˈgæf	756	bɪˈgæɸ	677	bɪˈgæp	711	began (50)
T58	bɪˈsaɪf	937	bɪˈsaɪɸ	969	bɪˈsɑɪk	748	beside (50)
T59	ˈbedrʊf	733	ˈbedrʊɸ	730	ˈbedrʊt	793	bedroom (35)
T60	ˈɪnpʊf	900	ˈɪnpʊɸ	796	ˈɪnpʊp	908	input (30)

A.2 Filler items

A.2.1 Monosyllabic pseudowords (n=30)

Code	stimulus	
FN01	ɔp	707
FN02	bɔɪt	669
FN03	klæt	862
FN04	dɔt	726
FN05	dæp	564
FN06	drɒk	770
FN07	glɑt	711
FN08	gæk	740
FN09	glɒp	532
FN10	haʊk	687
FN11	fɔp	927
FN12	praɪt	748
FN13	sɔk	763
FN14	jut	831
FN15	saɪp	975

Code	stimulus	
FN16	ænk	642
FN17	bəʊk	704
FN18	skɒp	671
FN19	skrʌk	871
FN20	ʃʊp	669
FN21	slɒk	952
FN22	smæt	744
FN23	truk	763
FN24	splæk	774
FN25	saʊnk	870
FN26	frɒp	665
FN27	θræk	1049
FN28	kləʊt	729
FN29	wʊt	651
FN30	glɑp	664

A.2.2 Disyllabic pseudowords (n=30)

Code	stimulus	
FN31	ˈædʌlk	737
FN32	ˈbækdrɒt	934
FN33	bɪˈkɒt	861
FN34	ˈɔfʊp	900
FN35	ˈdjulæt	991
FN36	ˈbedrɒp	892
FN37	enˈtræt	1128
FN38	ˈfɔtɒk	1084
FN39	əˈmʌt	964
FN40	ˈgʊmdrɒk	907
FN41	mˈʌk	803
FN42	rɪˈkɔp	879
FN43	ˈlaɪtʃɪk	1143
FN44	ˈmɪlksɒt	1066
FN45	ˈaʊtkrɒk	1003

Code	stimulus	
FN46	ˈsʌmtaɪk	932
FN47	səˈpəʊt	1214
FN48	ˈstɒpgæk	862
FN49	ˈtɪkʌt	1003
FN50	ˈwɪndəuk	1023
FN51	ədˈmɪp	832
FN52	əˈbrɔk	1071
FN53	ˈkætgʌp	775
FN54	ˈkɑləʊp	910
FN55	ˈʧekʊt	912
FN56	ˈɪgɒp	734
FN57	ˈlɒknʌp	1167
FN58	ˈmæskɒp	1030
FN59	ˈdeɪbʊt	752
FN60	əˈlɒp	643

A.2.3 Repeated monsyllabic words (n=20)

Code	f stimulus	
W01	bluff	476
W02	cough	477
W03	snuff	841
W04	dwarf	746
W05	fife	723
W06	gaff	495
W07	golf	517
W08	graph	725
W09	half	590
W10	hoof	631

Code	non-f stimulus	
W11	blot	650
W12	brook	583
W13	cat	555
W14	dart	795
W15	group	550
W16	float	710
W17	halt	654
W18	pike	664
W19	point	726
W20	swap	679

A.2.4 Repeated disyllabic words (n=20)

Code	f stimulus	
W21	aloof	775
W22	behalf	707
W23	carafe	892
W24	take-off	866
W25	dandruff	860
W26	enough	829
W27	sunroof	1011
W28	foolproof	1181
W29	giraffe	912
W30	fishwife	1134

Code	non-f stimulus	
W31	amok	724
W32	outlook	845
W33	impart	991
W34	discount	965
W35	make-up	798
W36	boycott	449
W37	offshoot	977
W38	playwright	909
W39	tightrope	851
W40	report	990

A.2.5 Non-repeated monosyllabic words (n=30)

Code	f stimulus	
FW01	cuff	566
FW02	spoof	684
FW03	staff	806
FW04	laugh	589
FW05	knife	647
FW06	toff	587
FW07	rough	598
FW08	trough	644
FW09	roof	513
FW10	wharf	477

Code	non-f stimulus	
FW11	honk	773
FW12	slope	1105
FW13	joint	800
FW14	lap	894
FW15	mask	973
FW16	type	635
FW17	nook	898
FW18	naught	995
FW19	scout	803
FW20	trot	580
FW21	ark	965
FW22	block	585
FW23	boat	812
FW24	cap	833
FW25	drop	579
FW26	flute	982
FW27	grope	594
FW28	look	555
FW29	oak	595
FW30	slight	945

A.2.6 Non-repeated disyllabic words (n=30)

Code	f stimulus	
FW31	handcuff	918
FW32	rebuff	763
FW33	reproof	906
FW34	riffraff	712
FW35	seraph	761
FW36	dyestuff	1066
FW37	show-off	828
FW38	soundproof	1116
FW39	housewive	1004
FW40	tip-off	665

Code	non-f stimulus	
FW41	devout	825
FW42	mistook	784
FW43	padlock	1255
FW44	bagpipe	1074
FW45	chestnut	1062
FW46	compote	937
FW47	dilute	1045
FW48	unhook	864
FW49	airport	877
FW50	hold-up	1226
FW51	acute	905
FW52	banknote	1085
FW53	claptrap	871
FW54	drawback	1331
FW55	exploit	1459
FW56	fortnight	1128
FW57	grown-up	763
FW58	postmark	1369
FW59	seaport	1423
FW60	turnpike	1309

A.2.7 Replacement stimuli for monitoring task

Code	pseudoword stimulus	
FN01	ɔps	790
FN03	klæts	807
FN06	drɒks	750
FN08	gæks	768
FN11	fɔps	1024
FN15	saɪps	1111
FN17	bəʊks	706
FN21	slɒks	921
FN27	θræks	1017
FN28	kləʊts	761
FN33	bɪˈkɒts	896
FN36	ˈbedrɒps	881
FN39	əˈmʌts	971
FN43	ˈlaɪtʃɪks	1157
FN45	ˈaʊtrɒks	1066
FN47	səˈpəʊts	1043
FN49	ˈtikʌts	1004
FN52	əˈbrɔks	1019
FN56	ˈlɒknʌps	788
FN58	ˈmæskɒps	1205

Code	word stimulus	
FW12	slopes	946
FW14	laps	817
FW15	masks	907
FW17	nooks	807
FW18	naughts	853
FW21	arks	741
FW23	boats	812
FW24	caps	840
FW26	flutes	909
FW28	looks	910
FW43	padlocks	968
FW44	bagpipes	1098
FW45	chestnuts	1072
FW50	hold-ups	1067
FW52	banknotes	1120
FW54	drawbacks	1211
FW55	exploits	1273
FW58	postmarks	1329
FW59	seaports	1386
FW60	turnpikes	1362

B/ Perl scripts

B.1 Script to generate the stimulus lists

```
#!/usr/bin/perl

# stim_list_monitoring
#
# Creates a stimulus list for the repetition priming task
############################################################


# Subroutines
#

# PLACE must be passed a list of stimuli.
# It places the stimuli randomly into the empty fields
# of @stim_list.
# 1 is added because of the e-prime header line.
#
sub place ($$$) {
    my $type = pop @_;
    my $resp = pop @_;
    my @list = shuffle(@_);
    foreach (@list) {
my @fields = split(/:/);
my $pos = 1 + int rand 360;
while ($stim_list[$pos]) {
     $pos = 1 + int rand 360;
}
$stim_list[$pos] = "$pos\t1\t\tTestProc\t$fields[0]\t$resp\t$type
\tNULL\t$fields[1]\t$fields[2]\t$fields[3]";
    }
}


# PLACE_PAIRS must be passed a list of stimuli pairs and a number.
# It places the stimuli randomly into the empty fields of
# @stim_list, the distance between pairs determined by the number.
# 1 is added because of the e-prime header line.
```

```
#
sub place_pairs ($$$$) {
    my $type = pop @_;
    my $resp = pop @_;
    my $dist = pop @_;
    my @list = @_;
    foreach (@list) {
my @fields = split(/:/);
my $pos1 = 1 + int rand 360-$dist;
my $pos2 = $pos1+$dist;
while ($stim_list[$pos1] || $stim_list[$pos2]) {
     $pos1 = 1 + int rand 360-$dist;
     $pos2 = $pos1+$dist;
}
my $r = int rand 2;
if ($r == 0) {
     $stim_list[$pos1] = "$pos1\t1\t\tTestProc\t$fields[0]\t$resp\t
$type\t$dist\t$fields[1]\t$fields[2]\t$fields[3]";
     $stim_list[$pos2] = "$pos2\t1\t\tTestProc\t$fields[4]\t$resp\t
$type\t$dist\t$fields[5]\t$fields[6]\t$fields[7]";
} else {
     $stim_list[$pos1] = "$pos1\t1\t\tTestProc\t$fields[4]\t$resp\t
$type\t$dist\t$fields[5]\t$fields[6]\t$fields[7]";
     $stim_list[$pos2] = "$pos2\t1\t\tTestProc\t$fields[0]\
$resp\t$type\t$dist\t$fields[1]\t$fields[2]\t$fields[3]";
}
    }
}


# PLACE_PAIR_LIST takes care of stimuli lists containing pairs.
# It must be passed a list containing pairs and it places the
# stimuli at distances of 8, 10, 12 and 14 fields in @stim_list
# using the subroutine PLACE_PAIRS.
#
sub place_pair_list ($$$) {
    my $type = pop @_;
    my $resp = pop @_;
    my @shuff = shuffle(@_);
    my $n = @shuff/4;
    my @list8 = @shuff[0..$n-1]; $dist8 = 8;
    my @list10 = @shuff[$n..2*$n-1]; $dist10 = 10;
    my @list12 = @shuff[2*$n..3*$n-1]; $dist12 = 12;
    my @list14 = @shuff[3*$n..4*$n-1]; $dist14 = 14;
    &place_pairs(@list8, $dist8, $resp, $type);
    &place_pairs(@list10, $dist10, $resp, $type);
    &place_pairs(@list12, $dist12, $resp, $type);
    &place_pairs(@list14, $dist14, $resp, $type);
}


# PLACE_TEST_LIST places the stimuli from the test list, making
```

```
# sure that the right stimuli are paired up for the 3 categories
# IDENTICAL, RELATED and UNRELATED.  It uses the subroutine
# PLACE_PAIR_LIST to do this.
#
sub place_test_list ($$$) {
    my $type = pop @_;
    my $resp = pop @_;
    my @shuff = shuffle(@_);
    my $n = @shuff/3;
    my @identical;
    my @related;
    my @unrelated;
    foreach (@shuff[0..$n-1]) {
my @fields = split(/:/);
push(@identical,"$fields[0]:$fields[1]:$fields[2]:$fields[3]:
$fields[0]:$fields[1]:$fields[2]:$fields[3]");
}
foreach (@shuff[$n..2*$n-1]) {
my @fields = split(/:/);
push(@related,"$fields[0]:$fields[1]:$fields[2]:$fields[3]:
$fields[4]:$fields[5]:$fields[6]:$fields[7]");
    }
foreach ( @shuff[2*$n..3*$n-1]) {
my @fields = split(/:/);
push(@unrelated,"$fields[0]:$fields[1]:$fields[2]:$fields[3]:
$fields[8]:$fields[9]:$fields[10]:$fields[11]");
    }
    &place_pair_list(@identical, $resp, $type);
    &place_pair_list(@related, $resp, $type);
    &place_pair_list(@unrelated, $resp, $type);
}


# Body of script
#
#####################################################################

use List::Util 'shuffle';
srand;

# This block reads in all the files into corresponding arrays
open IN, "input/test3" or die "Cannot read input/test3: $!";
chomp(@test = <IN>);
open IN, "input/words_rep3" or die "Cannot read input/words_rep3: $!";
chomp(@words_rep = <IN>);
open IN, "input/words3" or die "Cannot read input/words3: $!";
chomp(@words = <IN>);
open IN, "input/nonwords3" or die "Cannot read input/nonwords3: $!";
chomp(@nonwords = <IN>);
close IN;

# choose number of stimulus lists to be generated
```

```
for (1..80) {

# Creat empty list
    @stim_list = ();

# Add e-prime header
    $stim_list[0] = "ID\tWeight\tNested\tProcedure\tStimulus\t
    CorrectResp\tStimulusType\tDistance\tAlignmentPoint\tEndPoint
    \tMonitoring";


# Placement of learning stimuli (fixed place for all lists)
    $stim_list[4] =  "4\t1\t\tTestProc\tlf1.wav\t1\tlearn\t10\t558
\t807\t0";
    $stim_list[14] = "14\t1\t\tTestProc\tlp1.wav\t1\tlearn\t10
\t558\t807\t0";
    $stim_list[19] = "19\t1\t\tTestProc\tlp2.wav\t1\tlearn\t10
\t605\t854\t0";
    $stim_list[29] = "29\t1\t\tTestProc\tlf2.wav\t1\tlearn\t10
\t605\t854\t0";
    $stim_list[34] = "34\t1\t\tTestProc\tlf1.wav\t1\tflearn\tNULL
\t558\t807\t0";
    $stim_list[42] = "42\t1\t\tTestProc\tlp2.wav\t1\tflearn\tNULL
\t605\t854\t0";
    $stim_list[50] = "50\t1\t\tTestProc\tlp1.wav\t1\tflearn\tNULL
\t558\t807\t0";
    $stim_list[58] = "58\t1\t\tTestProc\tlf2.wav\t1\tflearn\tNULL
\t605\t854\t0";
    $stim_list[68] = "68\t1\t\tTestProc\tlf2.wav\t1\tflearn\tNULL
\t605\t854\t0";
    $stim_list[76] = "76\t1\t\tTestProc\tlp1.wav\t1\tflearn\tNULL
\t558\t807\t0";
    $stim_list[85] = "85\t1\t\tTestProc\tlp2.wav\t1\tflearn\tNULL
\t605\t854\t0";
    $stim_list[94] = "94\t1\t\tTestProc\tlf1.wav\t1\tflearn\tNULL
\t558\t807\t0";
    $stim_list[105] = "105\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[113] = "113\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[123] = "123\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[131] = "131\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[141] = "141\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[150] = "150\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[158] = "158\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[169] = "169\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
```

```
    $stim_list[178] = "178\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[186] = "186\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[195] = "195\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[206] = "206\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[214] = "214\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[224] = "224\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[232] = "232\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[243] = "243\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[253] = "253\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[261] = "261\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[270] = "270\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[278] = "278\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[287] = "287\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[298] = "298\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[307] = "307\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[315] = "315\t1\t\tTestProc\tlf2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[324] = "324\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[334] = "334\t1\t\tTestProc\tlf1.wav\t1\tflearn
\tNULL\t558\t807\t0";
    $stim_list[345] = "345\t1\t\tTestProc\tlp2.wav\t1\tflearn
\tNULL\t605\t854\t0";
    $stim_list[354] = "354\t1\t\tTestProc\tlp1.wav\t1\tflearn
\tNULL\t558\t807\t0";


# Placement of other stimuli
# 1 = word reponse; 5 = nonword response (SR Box!)
    &place_test_list(@test,5,test);
    &place_pair_list(@words_rep,1,word);
    &place(@words,1,fword);
    &place(@nonwords,5,fnon);


# Output to file
    open OUT, ">output/stimulus_list$_.txt"
```

```
or die "Cannot create output/stimulus_list$_.txt: $1";
    for ($i = 0; $i <= $#stim_list-1; $i++) {
print OUT "$stim_list[$i]\n";
    }
    for ($i = $#stim_list; $i <= $#stim_list+1; $i++) {
print OUT "$stim_list[$i]";
    }
    close OUT;
}
```

B.2 Script to transform reaction time to priming

```
#!/usr/bin/perl

# RT_difference5
#
# For procedure where presses and monitoring responses, but not
# releases, are logged.
# This computes RT-differences from the beginning, alignment point
# and end.
#


# check number of arguments
if (@ARGV != 1) {
    die "Usage: RT_difference FILENAME.\n";
}


# Array for output
@output = "Subject\tStimulusType\tPrimingRelationship\tStimulus
\tPrime\tProbe\tPrimePos\tDistance\tPrimeResp\tProbeResp
\tPress.RTdiff\tPress.RTdiff.aligned\tPress.RTdiff.end";


# Get filename for output
$name = "Priming$1" if ($ARGV[0] =~ /^RT(.*)/);


# Open filehandle and read content of file into an array
# (line by line)
open IN, "$ARGV[0]"
    or die "Cannot read $ARGV[0]: $!";
@input = <IN> while (<IN>);
close IN;


# Collection of data
for ($i = 0; $i <= $#input; $i++) {
    my $line = $input[$i];
```

```
# Check whether line is of right type
    if ($line =~ /\t(test|word)\t/) {
my @fields1 = split(/\t/, $line);
my $probe = $input[$i+$fields1[4]]; # $probe = line of probe
# print "$probe\n";

# Store name of prime
if ($fields1[3] =~ /((t|w)[pf]?([0-9][0-9]?))\.wav/) {
    my $stimulus = $1; # $stimulus contains name of prime
    # print "$stimulus\n";


# Check whether probe stimulus corresponds to prime (and subjects
# are identical)
    if ($probe =~ /(^$fields1[0].*\t$2[pf]?$3)/) {
my @fields2 = split(/\t/, $probe);

my $press_rt;
my $press_rt_aligned;
my $press_rt_end;


# Dealing with missing press RT data
if ($fields1[12] == "NA" || $fields2[12] == "NA" ||
$fields1[12] == "NULL" || $fields2[12] == "NULL") {
    $press_rt = "NA";
    $press_rt_aligned = "NA";
    $press_rt_end = "NA";
} else {


# Press RT data: rt = RT difference; rt_aligned = RT difference
# from alignment point; rt_end RT difference from end
    $press_rt = $fields1[12] - $fields2[12];
    $press_rt_aligned = $fields1[14] - $fields2[14];
    $press_rt_end = $fields1[16] - $fields2[16];
}


# Code prime-probe relationship
my $priming;
if ($fields1[3] =~ /^(w|l|f)/) {
    $priming = "NA";
} elsif ($fields1[3] eq $fields2[3]) {
    $priming = "identical";
} elsif (($fields1[3] =~ /^tf/ && $fields2[3] =~ /^tp/) ||
($fields1[3] =~ /^tp/ && $fields2[3] =~ /^tf/)) {
    $priming = "related";
} else {
    $priming = "unrelated";
}
```

```
# Type of test stimulus
my $type;
if ($fields1[3] =~ /t[a-z]?([0-9][0-9]?)\.wav/) {
    $type = $1;
} else {
    $type = "NA";
}


# Put all the information into the @output array
push (@output, "$fields1[0]\t$fields1[2]\t$priming\t$type
\t$fields1[3]\t$fields2[3]\t$fields1[1]\t$fields1[4]\t$fields1[7]
\t$fields2[7]\t$press_rt\t$press_rt_aligned\t$press_rt_end");
    }
}
    }
}


# Write to file
open OUT, ">$name"
    or die "Cannot create file '$name': $!";
foreach (@output) {
    print OUT "$_\n";
}
close OUT;
```

B.3 Script to generate the acoustic continua

```
#!/usr/bin/perl


# continuum
#
# This script takes two Praat Matrix-files and creates a
# continuum by merging the two files sample by sample in
# different proportions.
# This is a command-line script that works similar to a
# Unix command.
# It takes three arguments: the number of steps in the
# continuum (prefixed by -), and the two files to be merged.
# NB: The script does not check whether the construction of
# the contrast makes sense; this is up to the user to determine.
#


if (@ARGV != 3) {
    die "Usage: -STEPS FILE1 FILE2\n(STEPS = number of steps in
```

```
 continuum; FILE1 & FILE2 = Praat Matrix-files).\n";
}

# Compute distance from step number
if ($ARGV[0] =~ /-([0-9]+)/) {
    $n_steps = $1;
}

if ($n_steps != 0) {
    $distance = 1/($n_steps);
} else {
    $distance = 0;
}


# Open filehandles for input files and read content into two
#  arrays (line by line)
open IN_1, "$ARGV[1]"
    or die "Cannot read $ARGV[1]: $!";
while(<IN_1>) {
    @list1 = <IN_1>;
      }
close IN_1;

open IN_2, "$ARGV[2]"
    or die "Cannot read $ARGV[2]: $!";
while(<IN_2>) {
    @list2 = <IN_2>;
      }
close IN_2;


# Output file name (not used at the moment: see line 69)
if ($ARGV[1] =~ /^\w+\.Matrix$/ and $ARGV[2] =~ /^\w+\.Matrix$/) {
    $name1 = $1 if ($ARGV[1] =~ /^(\w+)\.Matrix$/);
    $name2 = $1 if ($ARGV[2] =~ /^(\w+)\.Matrix$/);
    $name = "${name1}_${name2}_$n_steps";
} else {
    $name = "merged_$n_steps"
    }


# Create array for head of output files (NB first line supplied
# because the first line of the input files fail to be read
@head = "File Type = \"ooTextFile\"\n";
foreach (2..15) {
    @head[$_] = shift(@list1);
    shift @list2;
}


# Create $n_steps number of files
```

```
for ($i=0; $i <= $n_steps; $i++){
    if ($i < 10) {
open OUT, ">0${i}.Matrix"
    or die "Cannot create file '$name_$i.Matrix': $!";
my @merged = @head;


# Process the input line by line
for ($j=0; $j< @list1; $j++){
    my $text; my $number1; my $number2;
   if ($list1[$j] =~ /(z \[[0-9+]\] \[[0-9]+\] = )(0|-?[0-9]+\.[0-9]+)
    \W*$/) {
$text = $1; $number1 = $2; # split number from text
    }
    if ($list2[$j] =~ /(0|-?[0-9]+\.[0-9]+)\W*$/) {
$number2 = $1; # ditto
    }
    $new_number = ($number1 * (1 - $i*$distance)) + ($number2
    * $i*$distance);
   $text = "\t" . $text . $new_number . "\n"; # compute output number
  push(@merged, $text);
}
print OUT @merged;
close OUT;
    } else {
open OUT, ">${i}.Matrix"
    or die "Cannot create file '$name_$i.Matrix': $!";
my @merged = @head;


# Process the input line by line
for ($j=0; $j< @list1; $j++){
    my $text; my $number1; my $number2;
   if ($list1[$j] =~ /(z \[[0-9+]\] \[[0-9]+\] = )(0|-?[0-9]+\.[0-9]+)
    \W*$/) {
$text = $1; $number1 = $2; # split number from text
    }
    if ($list2[$j] =~ /(0|-?[0-9]+\.[0-9]+)\W*$/) {
$number2 = $1; # ditto
    }
    $new_number = ($number1 * (1 - $i*$distance)) + ($number2
    * $i*$distance);
   $text = "\t" . $text . $new_number . "\n"; # compute output number
    push(@merged, $text);
}
print OUT @merged;
close OUT;
    }
}
```

C/ The informed consent form

Informed Consent Form

Please read the following information carefully. You can also request a copy.

Experiment: Word learning
Experimenter: Lukas Wiget
Affiliation: Linguistics and English Language, University of Edinburgh

Description

You are invited to participate in an experimental study that investigates the acquisition of new words. The experiment takes place over three sessions. There will be two training sessions, where you will learn four words that are used in a variety of English spoken in Sri Lanka to refer to local plants. In the third and final session, you will be tested on these and other words. Note that towards the end of the first session you will have to recognise the new words about 80% of times in order to progress to the second session.

Risks and benefits

There are no known risks involved in the experiment. There are no benefits to participation beyond the remuneration that you will receive.

Time involvement and payment

Sessions 1 and 2 will take about 30 minutes each, session 3 will take about 50 minutes to complete. You will receive £3 for the two shorter sessions and £5 for the third session, plus a bonus of £3 for completing the study. Payment will be made at the end of the third session.

Subject rights

Please understand that your participation is voluntary and that you have the right to withdraw your consent or discontinue participation at any time. Your privacy will be maintained in all published and written data resulting from the study.

If you agree with the conditions stated above and are willing to participate in the experiment, please sign below. By signing the form you confirm that you meet the following conditions:

- You are a native speaker of English.
- You are at least 18 years of age.
- You have no known hearing deficiencies.
- You have read the above consent form, understood it, and agree to it.
- You want to participate in the above-mentioned experiment.

Some additional questions:

- Do you speak any foreign languages? _____
- Which dialect of English do you speak? _____
- Are you right- or left-handed? _____
- Your age: _____

Name:

_____

Date: _____ Signature: _____

Bibliography

Abramson, A. S. and Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, pages 569–573. Academia, Prague.

Allopenna, P., Magnuson, J., and Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38:419–439.

Andruski, J., Blumstein, S. E., and Burton, M. W. (1994). The effet of subphonemic differences on lexical access. *Cognition*, 52:163–187.

Baayen, R. H. (2004). Statistics in psycholinguistics: a critique of the current gold standards. *Mental Lexicon Working Papers*, 1:1–45.

Baayen, R. H., Davidson, D., and Bates, D. M. (submitted). Mixed-effects modeling with crossed random effects for subjects and items. Manuscript submitted for publication.

Baayen, R. H., Piebenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 2nd edition.

Bard, E. G. and Shillcock, R. C. (1993). Competitor effects during lexical access: chasing Zipf's tail. In Altmann, G. T. and Shillcock, R. C., editors, *Cognitive models of speech processing: the second Sperlonga meeting*, pages 235–275. Lawrence Erlbaum, Hove.

Bates, D. M. and Sarkar, D. (2007). *lme4: linear mixed-effects models using S4 classes*. R Foundation for Statistical Computing.

Best, C. T., Morrongiello, B., and Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 29:191–211.

Blumstein, S. E., Milberg, W. P., Brown, T., Hutchinson, A., Kurowski, K., and Burton, M. W. (2000). The mapping from sound structure to the lexicon in aphasia: evidence from rhyme and repetition priming. *Brain and Language*, 72:75–99.

Blumstein, S. E. and Stevens, K. N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66:1001–1017.

Blumstein, S. E. and Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 67:648–662.

Boersma, P. and Weenink, D. (2006). *Praat: doing phonetics by computer*.

Bölte, J. and Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner? *Brain and Language*, 81:384–397.

Bölte, J. and Uhe, M. (2004). When is all understood and done? The psychological reality of the recognition point. *Brain and Language*, 88:133–147.

Bond, T. E. T. (1953). *Wild flowers of the Ceylon hills: some familiar plants of the up-country districts*. Oxford University Press, Madras.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49:155–180.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper and Row, New York.

Church, B. and Schacter, D. L. (1994). Perceptual specificity of auditory priming: memory for voice, intonation, and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:521–533.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12:335–359.

Clark, J. and Yallop, C. (1995). *An introduction to phonetics and phonology*. Blackwell, Oxford, 2nd edition.

Coady, J. A. and Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, 89:183–213.

Coleman, J. (1998). Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics*, 11:295–320.

Connine, C. M. (2004). It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, pages 1084–1089.

Connine, C. M., Blasko, D., and Titone, D. (1993). Do the beginnings of words have a special status in auditory word recognition? *Journal of Memory and Language*, 32:193–210.

Corina, D. (1992). Syllable priming and lexical representations: evidence from experiments and simulations. In *14th Annual Conference of the Cognitive Science Society*, pages 779–784, Bloomington, IN. Indiana University.

Craik, F. and Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 26:274–284.

Cutler, A. and Donselar, W. v. (2001). Voornam is not (really) a homophone: lexical prosody and lexical access in Dutch. *Language and Speech*, 44:171–195.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes*, 16:507–534.

Dahan, D. and Mead, R. (2006). The role of acoustic similarity in listeners' adaptation to artificially distorted speech. In *AMLaP 2006: the 12th Annual Conference on Architectures and Mechanisms for Language Processing*.

Davenport, J. and Dickinson, R. (1973). A comparison of some approximate F-tests. *Technometrics*, 15:779–789.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134:222–241.

Diehl, R. L. and Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psy-*

chology, 1:123–144.

Dufour, S. and Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: when the target's competitors conflict with the prime word. *Cognition*, 88:B33–B44.

Dumay, N., Benraïs, A., Barriol, B., Colin, C., and Radeau, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, 13:121–143.

Eimas, P. D. and Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4:99–109.

Eisner, F. and McQueen, J. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, 67:224–238.

Elman, J. L. and McClelland, J. L. (1984). Speech perception as a cognitive process: the interactive activation model. In Lass, N. J., editor, *Speech and language: advances in basic research and practice*, volume 10, pages 337–374. Academic Press, New York.

Emmorey, K. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4:73–92.

Estes, W. (1993). Concepts, categories, and psychological science. *Psychological Science*, 4:143–153.

Faraway, J. J. (2005). *Linear models with R*. Chapman & Hall/CRC, Boca Raton.

Faraway, J. J. (2006). *Extending the linear model with R*. Chapman & Hall/CRC, Boca Raton.

Feustel, T., Shiffrin, R., and Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, 112:309–346.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory and Cognition*, 5:335–339.

Fitch, H., Hawles, T., Erickson, D., and Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics*, 27:343–350.

Flege, J. E. (1992). Speech learning in a second language. In Ferguson, C. A. and Menn, L., editors, *Phonological development: models, research, implications*, pages 565–604. York Press, Timonium, MD.

Flege, J. E. and Hillenbrand, J. (1986). Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of english. *Journal of the Acoustical Society of America*, 79:508–517.

Forster, K. I. and Dickinson, R. (1976). More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F', and minF'. *Journal of Verbal Learning and Verbal Behavior*, 15:135–142.

Fowler, C. and Rosenblum, L. (1991). The perception of phonetic gestures. In Mattingly, I. and Studdert-Kennedy, M., editors, *Modularity and the motor theory of speech perception*, pages 33–59. Lawrence Erlbaum, Hillsdale, NJ.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Sage, Thousand Oaks.

Fox, R. (1984). Effects of lexical status on phonetic categorization. *Journal of Experimental*

Psychology: Human Perception and Performance, 10:526–540.

Frauenfelder, U. H., Scholten, M., and Content, A. (2001). Bottom-up inhibition in lexical selection: phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16:583–607.

Frisch, S. A., Large, N. R., and Pisoni, D. B. (2000). Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42:481–496.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5:171–189.

Fujisaki, H. and Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 28:67–73.

Fujisaki, H. and Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 29:207–214.

Ganong, D., Palmer, N., and Sawusch, J. R. (2001). The segmental representation of spoken words. In McLennan, C., Luce, P. A., Mauner, G., and Charles-Luce, J., editors, *University of Buffalo Working Papers on Language and Perception*, volume 1, pages 151–255.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6:110–125.

Gaskell, G. and Marslen-Wilson, W. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, 45:220–266.

Gaskell, M. G. (2000). Modeling lexical effects on phonetic categorization and semantic effects on word recognition. *Behavioral and Brain Sciences*, 23:329–339.

Gaskell, M. G. and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89:105–132.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., and Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25:84–95.

Gerrits, E. and Schouten, M. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics*, 66:363–376.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105:251–279.

Goldinger, S. D. (1999). Only the shadower knows: comment on Hamburger & Slowiaczek (1996). *Psychonomic Bulletin and Review*, 60:347–351.

Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: effects of competition and inhibition. *Journal of Memory and Language*, 28:501–518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., and Marcario, J. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 18:1211–1238.

Goldinger, S. D., Pisoni, D. B., and Luce, P. A. (1996). Speech perception and spoken word recognition: research and theory. In Lass, N. J., editor, *Principles of Experimental Phonetics*, pages 277–327. Mosby, St. Louis.

Greenberg, S. (1996). Auditory processing of speech. In Lass, N. J., editor, *Principles of experimental phonetics*, pages 362–407. Mosby, St. Louis.

Guenther, F. H. (2003). Neural control of speech movements. In Meyer, A. and Schiller, N., editors, *Phonetics and phonology in language comprehension and production: differences and similarities*, pages 209–240. Mouton de Gruyter, Berlin.

Halle, M. and Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report*, 101:198–212.

Hamburger, M. and Slowiaczek, L. M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin and Review*, 3:520–525.

Hamburger, M. and Slowiaczek, L. M. (1999). On the role of bias in dissociated phonological priming effects: a reply to Goldinger (1999). *Psychonomic Bulletin and Review*, 6:352–355.

Hayward, K. (2000). *Experimental phonetics*. Longman, Harlow.

Healy, A. F. and Repp, B. H. (1982). Context independence in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 8:68–80.

Hintzman, D. (1984). Minerva 2: a simulation model of human memory. *Behavior Research Methods, Instruments, and Computers*, 16:96–101.

Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95:528–551.

Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, 93:411–428.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*, 10:785–813.

Jacoby, L. (1983). Perceptual enhancement: persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9:21–38.

Johnson, K. (1997). Speech perception without speaker normalization. In Johnson, K. and Mullennix, J. W., editors, *Talker variability in speech processing*, pages 145–165. Academic Press, San Diego.

Jusczyk, P. W. (1993). From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21:3–28.

Jusczyk, P. W. (1997). *The discovery of spoken language*. MIT Press, Cambridge, MA.

Jusczyk, P. W. and Luce, P. A. (1994). Infants' sensitivity to phonotactic patterns in their native language. *Journal of Memory and Language*, 33:630–645.

Kirsner, K., Dunn, J. C., and Standen, P. (1987). Record-based word recognition. In Coltheart, M., editor, *Attention and performance XII: the psychology of reading*, pages 147–167. Lawrence Erlbaum, Hillsdale, NJ.

Klatt, D. H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7:279–312.

Klatt, D. H. (1989). Review of selected models of speech perception. In Marslen-Wilson, W.,

editor, *Lexical representation and process*, pages 196–226. MIT Press, Cambridge, MA.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In Gernsbacher, M. A., editor, *Handbook of psycholinguistics*, pages 173–217. Academic Press, New York.

Kruschke, J. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, 99:22–44.

Kuhl, P. K. and Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63:905–917.

Ladefoged, P. and Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell, Oxford.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press, Cambridge.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, 3:223–232.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74:431–461.

Liberman, A. M., Harris, K. S., Hoffman, H., and Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54:358–368.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21:1–36.

Lipinski, J. and Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language*, 52:171–192.

Lisker, L. and Abramson, A. S. (1970). The voicing dimension: some experiments in comparative phonetics. In Hála, B., Romportl, M., and Janota, P., editors, *Proceedings of the Sixth International Congress of Phonetic Sciences*, pages 563–568.

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin and Review*, 7:618–630.

Luce, P. A. (1986). *Neighbourhoods of words in the mental lexicon*. Doctoral dissertation, Indiana University.

Luce, P. A., Goldinger, S. D., Auer, E. T., and Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, 62:615–625.

Luce, P. A. and Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16:565–581.

Luce, P. A. and Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, 26:708–715.

Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In Altmann, G. T., editor, *Cognitive models of speech perception: psycholinguistics and computational perspectives*, pages 122–147. MIT Press, Cambridge, MA.

Luce, R. (1959). *Individual choice behavior*. Wiley, New York.

Macmillan, N. A. (1987). Beyond the categorical/continuous distinction: a psychophysical approach to processing modes. In Harnad, S., editor, *Categorical perception: the groundwork of cognition*, pages 53–85. Cambridge University Press, Cambridge.

Macmillan, N. A. and Creelman, C. D. (2005). *Detection theory: a user's guide*. Lawrence

Erlbaum Associates, Mahwah, NJ, 2nd edition.

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, 84:452–471.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, Cambridge.

Magnuson, J., Tanenhaus, M., Aslin, R. N., and Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132:202–227.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25:71–102.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In Altmann, G. T., editor, *Cognitive models of speech processing: psycholinguistic and computational perspectives*, pages 148–172. MIT Press, Cambridge, MA.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In Altmann, G. T. and Shillcock, R. C., editors, *Cognitive models of speech processing: the second Sperlonga meeting*, pages 187–210. Lawrence Erlbaum, Hove.

Marslen-Wilson, W., Moss, H. E., and van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22:1376–1392.

Marslen-Wilson, W. and Tyler, L. K. (1980). The temporal structure of spoken word understanding. *Cognition*, 8:1–71.

Marslen-Wilson, W. and Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, 101:653–675.

Marslen-Wilson, W. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10:29–63.

Marslen-Wilson, W. and Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15:576–586.

Massaro, D. W. (1987). Categorical partition: a fuzzy-logical model of categorization behavior. In Harnad, S., editor, *Categorical perception: the groundwork of cognition*, pages 254–283. Cambridge University Press, Cambridge.

Maxwell, S. E. and Delaney, H. D. (2004). *Designing experiments and analyzing data: a model comparison perspective*. Lawrence Erlbaum, Mahwah, NJ.

May, J. G. (1981). Acoustic factors that may contribute to categorical perception. *Language and Speech*, 24:273–284.

Mayo, C. and Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America*, 115:3184–3194.

McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *American Statistician*, 32:12–16.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:539–553.

McQueen, J. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17:433–443.

McQueen, J. (1996). Phonetic categorization. *Language and Cognitive Processes*, 11:655–664.

McQueen, J., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30:1113–1126.

McQueen, J. and Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. In *ISCA Workshop on Plasticity in Speech Perception*, pages 233–236, London.

McQueen, J., Norris, D., and Cutler, A. (1999). Lexical influences in phonetic decision making: evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25:1363–1389.

McQueen, J. and Sereno, J. (2005). Cleaving automatic processes from strategic biases in phonological priming. *Memory and Cognition*, 33:1185–1209.

Medin, D. L. and Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, 85:207–238.

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90:227–234.

Miller, J. L., Eimas, P. D., and Zatorre, R. (1979). Studies of place and manner of articulation in syllable-final position. *Journal of the Acoustical Society of America*, 66:1207–1210.

Monsell, S. (1985). Repetition and the lexicon. In Ellis, A. O., editor, *Progress in the psychology of language*, volume II, pages 147–195. Lawrence Erlbaum Associates, London.

Monsell, S. and Hirsh, K. W. (1998). Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:1495–1520.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76:165–178.

Morton, J. (1979). Word recognition. In Morton, J. and Marshall, J., editors, *Psycholinguistics 2: structures and processes*, pages 107–156. MIT Press, Cambridge, MA.

Nearey, T. M. (1990). The segement as a unit of speech perception. *Journal of Phonetics*, 18:347–373.

Neely, J. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106:226–254.

Newman, R. S., Sawusch, J. R., and Luce, P. A. (1997). Lexical neighborhood effect in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 23:873–889.

Newman, R. S., Sawusch, J. R., and Luce, P. A. (2005). Do postonset segments define a lexical neighborhood? *Memory and Cognition*, 33:941–960.

Nittrouer, S. and Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30:319–329.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52:189–234.

Norris, D., McQueen, J., and Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:1209–1228.

Norris, D., McQueen, J., and Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, 23:299–325.

Norris, D., McQueen, J., and Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory and Cognition*, 30:399–411.

Norris, D., McQueen, J., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47:204–238.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14:700–708.

Oden, G. and Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85:172–191.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, 99(3):1718–1725.

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: concrete exemplar-based vs. abstract lexical entries. *Psychological Science*, 12:445–449.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19:309–328.

Pastore, R. E. (1987). Categorical perception: some psychophysical models. In Harnad, S., editor, *Categorical perception: the groundwork of cognition*, pages 29–52. Cambridge University Press, Cambridge.

Pastore, R. E., Ahroon, W. A., Buffuto, K. J., Friedman, C., Puelo, J. S., and Fink, E. A. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 3:686–696.

Patterson, D. and Connine, C. M. (2001). Variant frequency in flap production: a corpus analysis of variant frequency in American English flap production. *Phonetica*, 58:299–325.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and the acquisition of phonology. *Language and Speech*, 46:115–154.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer Verlag, New York.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13:253–260.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, 3:7–18.

Pisoni, D. B., Aslin, R. N., Perey, A., and Hennessy, B. (1982). Some effects of laboratory training on the identification and discrimination of voicing contrasts in stop consonants. *Journal of*

Experimental Psychology: Human Perception and Performance, 8:297–314.

Pisoni, D. B. and Luce, P. A. (1987). Acoustic-phonetic representation in word recognition. *Cognition*, 25:21–52.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:1037–1052.

Pitt, M. A. and Shoaf, L. (2002). Revisiting bias effects in word-initial phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, 28:1120–1130.

Polka, L. and Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, 100:577–592.

R Development Core Team (2006). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Raaijmakers, J., Schrijnemakkers, J., and Gremmen, F. (1999). How to deal with 'the language-as-fixed-effect fallacy': common misconceptions and alternative solutions. *Journal of Memory and Language*, 41:416–426.

Radeau, M., Morais, J., and Dewier, A. (1989). Phonological priming in spoken word recognition: task effects. *Memory and Cognition*, 17:525–535.

Radeau, M., Morais, J., and Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 21:1297–1311.

Rapahel, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 51:1296–1303.

Repp, B. H. (1981a). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 18:12–14.

Repp, B. H. (1981b). Two strategies in fricative discrimintion. *Perception and Psychophysics*, 30:217–227.

Repp, B. H. (1984). Categorical perception: issues, methods, findings. In Lass, N. J., editor, *Speech and language: advances in basic research and practice*, volume 10, pages 244–322. Academic Press, New York.

Rouder, J. N., Lu, J., Speckman, P., Sun, D., and Yi, J. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin and Review*, 12:195–223.

Sawusch, J. R. and Jusczyk, P. W. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance*, 7:408–421.

Schacter, D. L. (1992). Understanding implicit memory: a cognitive neuroscience approach. *American Psychologist*, 47:559–569.

Schacter, D. L. and Church, B. (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18:915–930.

Schouten, B., Gerrits, E., and van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41:71–80.

Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The influence of initial exposure on

lexical representation: comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52:240–255.

Sebastián-Gallés, N., Rodríguez-Fornells, de Diego-Balaguer, R., and Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, 18:1277–1291.

Seidenberg, M., Waters, G., Sanders, M., and Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition*, 12:315–328.

Shillcock, R. C. (1990). Lexical hypothesis in continuous speech. In Altmann, G. T., editor, *Cognitive models of speech processing: psycholinguistic and computational perspectives*. MIT Press, Cambridge, MA.

Slowiaczek, L. M. and Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18:1239–1250.

Slowiaczek, L. M., McQueen, J., Soltano, E. G., and Lynch, M. (2000). Phonological representations in prelexical speech processing: evidence from form-based priming. *Journal of Memory and Language*, 43:530–560.

Slowiaczek, L. M., Nusbaum, H. C., and Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13:64–75.

Slowiaczek, L. M. and Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition*, 14:230–237.

Snijders, T. A. and Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage, London.

Spinelli, E., Segui, J., and Radeau, M. (2001). Phonological priming in spoken word recognition with bisyllabic targets. *Language and Cognitive Processes*, 16:367–392.

Stevens, K. N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: data from phonology, acoustics, and psychoacoustics. In Myers, T., Laver, J., and Anderson, J., editors, *The cognitive representation of speech*, pages 61–74. North Holland, Amsterdam.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–45.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111:1872–1891.

Stevens, K. N. and Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64:1358–1368.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. E. G. (1969). Cross-language study of vowel percpetion. *Language and Speech*, 12:1–23.

Stork, H. L. (2001). Learning new words: phontactic probability in language development. *Journal of Speech and Hearing Research*, 44:1321–1337.

Streeter, L. and Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, 65:1533–1541.

Studdert-Kennedy, M. (1974). The perception of speech. In Sebeok, T., editor, *Current trends in linguistics*, volume 12, pages 2349–2385. Mouton, Den Haag.

Studdert-Kennedy, M. (1976). Speech perception. In Lass, N. J., editor, *Contemporary issues in experimental phonetics*, pages 243–293. Academic Press, New York.

Summerfield, Q. and Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 62:436–448.

Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes*, 11:569–576.

Taft, M. and Hambly, G. (1986). Exploring the Cohort Model of spoken word recognition. *Cognition*, 22:259–282.

Tanenhaus, M. K., Magnuson, J. S., McMurray, B., and Aslin, R. N. (2000). No compelling evidence against feedback in spoken word recognition. *Behavioral and Brain Sciences*, 23:348–349.

Tenpenny, P. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, 2:339–363.

Trask, R. (1996). *A dictionary of phonetics and phonology*. Routledge, London.

Treiman, R., Kessler, B., Kneewasser, S., Tincoff, R., and Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. In Broe, M. B. and Pierrehumbert, J. B., editors, *Papers in Laboratory Phonology V: acquisition and the lexicon*, pages 296–283. Cambridge University Press, Cambridge.

Tulving, E. (1972). Episodic and semantic memory. In Tulving, E. and Donaldson, W., editors, *Organization of memory*, pages 382–402. Academic Press, New York.

Vitevitch, M. S. and Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science*, 9:325–329.

Vitevitch, M. S. and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40:374–408.

Vitevitch, M. S. and Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, 52:193–204.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech*, 40(47–62).

Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception and Psychophysics*, 35:49–64.

Wheeldon, L. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken word recognition. *Cognition*, 32:25–64.